

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 31, 2004, 13:52:29 ; Search time 3483 Seconds  
(without alignments)  
7167.840 Million cell updates/sec

Title: US-09-975-856-1

Perfect score: 576

Sequence: 1 ATGAACGGACGACGACGCTT.....ATGACGAGTAACGCCCTCG 576

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 3470272 seqs, 21671516995 residues

Total number of hits satisfying chosen parameters: 6940544

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl.\*

1: gb\_ba.\*

2: gb\_htg.\*

3: gb\_in.\*

4: gb\_om.\*

5: gb\_ov.\*

6: gb\_pat.\*

7: gb\_ph.\*

8: gb\_pl.\*

9: gb\_pr.\*

10: gb\_ro.\*

11: gb\_sts.\*

12: gb\_sy.\*

13: gb\_un.\*

14: gb\_vi.\*

15: em\_ba.\*

16: em\_fun.\*

17: em\_hum.\*

18: em\_in.\*

19: em\_mu.\*

20: em\_om.\*

21: em\_or.\*

22: em\_ov.\*

23: em\_pat.\*

24: em\_ph.\*

25: em\_pl.\*

26: em\_ro.\*

27: em\_sts.\*

28: em\_un.\*

29: em\_vi.\*

30: em\_htg\_hum.\*

31: em\_htg\_inv.\*

32: em\_htg\_other.\*

33: em\_htg\_mus.\*

34: em\_htg\_pln.\*

35: em\_htg\_rnd.\*

36: em\_htg\_mam.\*

37: em\_htg\_vrt.\*

38: em\_sy.\*

39: em\_htgo\_hum.\*

40: em\_htgo\_mus.\*

41: em\_htgo\_other.\*

score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	576	100.0	576	6	AR167456	AR167456 Sequence
2	576	100.0	576	6	AR170503	AR170503 Sequence
3	576	100.0	576	6	AR182977	AR182977 Sequence
4	576	100.0	576	6	AR305671	AR305671 Sequence
5	576	100.0	576	6	AX740218	AX740218 Sequence
6	576	100.0	576	6	BD096872	BD096872 Isolated
7	576	100.0	576	6	BD218420	BD218420 SSX gene,
8	576	100.0	576	9	HSU90841	U90841 Homo sapien
9	574.4	99.7	1250	9	BC005325	BC005325 Homo sapi
10	496.6	86.2	766	6	A48452	A48452 Sequence 17
11	496.6	86.2	766	6	AX821947	AX821947 Sequence
12	496.6	86.2	766	9	HSSX2	X86175 H sapiens m
13	496.6	86.2	931	6	AR025465	AR025465 Sequence
14	496.6	86.2	931	6	AR060380	AR060380 Sequence
15	496.6	86.2	931	6	AR117874	AR117874 Sequence
16	496.6	86.2	931	6	AR167453	AR167453 Sequence
17	496.6	86.2	931	6	I85576	I85576 Sequence 2
18	496.6	86.2	931	6	AR287596	AR287596 Sequence
19	496.6	86.2	931	6	AR305668	AR305668 Sequence
20	496.6	86.2	931	6	AX114023	AX114023 Sequence
21	496.6	86.2	931	6	AX719099	AX719099 Sequence
22	496.6	86.2	931	6	BD218417	BD218417 SSX gene,
23	496.6	86.2	1309	6	AX331558	AX331558 Sequence
24	496.6	86.2	1309	6	AX331950	AX331950 Sequence
25	496.6	86.2	1309	9	HSHD21LEK	Z49105 H sapiens H
26	496	86.1	1249	9	HSU90840	U90840 Homo sapien
27	495	85.9	1276	9	BC016957	BC016957 Homo sapi
28	493.4	85.7	1282	9	BC007343	BC007343 Homo sapi
29	492.8	85.6	675	9	S82471	S82471 Homo sapien
30	482.2	83.7	576	6	AR167457	AR167457 Sequence
31	482.2	83.7	576	6	AR170504	AR170504 Sequence
32	482.2	83.7	576	6	AR182978	AR182978 Sequence
33	482.2	83.7	576	6	AR305672	AR305672 Sequence
34	482.2	83.7	576	6	AX740219	AX740219 Sequence
35	482.2	83.7	576	6	BD096873	BD096873 Isolated
36	482.2	83.7	576	6	BD218421	BD218421 SSX gene,
37	481.6	83.6	1261	9	BC001003	BC001003 Homo sapi
38	481.2	83.5	766	6	A48450	A48450 Sequence 15
39	480.6	83.4	576	9	HSU90842	U90842 Homo sapien
40	480	83.3	766	6	AR167452	AR167452 Sequence
41	480	83.3	766	6	AR305667	AR305667 Sequence
42	480	83.3	766	6	AX821946	AX821946 Sequence
43	480	83.3	766	6	BD218416	BD218416 SSX gene,
44	480	83.3	766	9	HSSX1	X86174 H sapiens m
45	433.2	75.2	1363	9	BC016640	BC016640 Homo sapi

# ALIGNMENTS

RESULT 1  
AR167456

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

AR167456 Sequence 5 from patent US 6287756. 576 bp DNA linear PAT 17-DEC-2001

AR167456

AR167456.1 GI:17903237

Unknown.

Unknown.

Unclassified.

1 (bases 1 to 576)

Tureci, O., Chen, Y.-T., Sahin, U., Gure, A.O., Old, L.J. and

Pfreundschuh, M.

Methods for determining presence of cancer in a sample by

determining expression of an SSX gene

Pred. No. is the number of results predicted by chance to have a

JOURNAL Patent: US 6287756-A 5 11-SEP-2001;  
 FEATURES Location/Qualifiers  
 source 1..576  
 /organism="unknown"  
 /mol\_type="unassigned DNA"

ORIGIN

Query Match 100.0%; Score 576; DB 6; Length 576;  
 Best Local Similarity 100.0%; Pred. No. 2.6e-164;  
 Matches 576; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ATGAACGGAGACGCGCTTTGCAAGGAGACCCAGGAGTGTCTCAAAATATCAGAGAAG 60  
 Db 1 ATGAACGGAGACGCGCTTTGCAAGGAGACCCAGGAGTGTCTCAAAATATCAGAGAAG 60  
 Qy 61 TTACGAAAGGCGCTTCGATGATATGCGCAAAATCTCTTAAGAAAGAGTGGGAAAAGATG 120  
 Db 61 TTACGAAAGGCGCTTCGATGATATGCGCAAAATCTCTTAAGAAAGAGTGGGAAAAGATG 120  
 Qy 121 AAATCCTCGGAGAAAATCGTCTATGTATATGAGCTAAATCTATGAGGTCTATGACTAAA 180  
 Db 121 AAATCCTCGGAGAAAATCGTCTATGTATATGAGCTAAATCTATGAGGTCTATGACTAAA 180  
 Qy 181 CTAGGTTTCAGGTCACCGCTCCACCTTTTCATGCGTAGTAAACGGGCTGCAGACTTCCAC 240  
 Db 181 CTAGGTTTCAGGTCACCGCTCCACCTTTTCATGCGTAGTAAACGGGCTGCAGACTTCCAC 240  
 Qy 241 GGGAAATGATTTTGGTAACGATCGAAACCAACAGGAAATCAGGTTGAACGTCCTCAGATGACT 300  
 Db 241 GGGAAATGATTTTGGTAACGATCGAAACCAACAGGAAATCAGGTTGAACGTCCTCAGATGACT 300  
 Qy 301 TTGCGACGCTCCAGAGAAATCTTCCGAAAGATCATGCGCAAGAAATGATGGGAAACAGCTGTGC 420  
 Db 301 TTGCGACGCTCCAGAGAAATCTTCCGAAAGATCATGCGCAAGAAATGATGGGAAACAGCTGTGC 420  
 Qy 361 AATGTTTGAAGGAGTCCAGAGGATCTCGCCACAAATGATGGGAAACAGCTGTGC 480  
 Db 361 AATGTTTGAAGGAGTCCAGAGGATCTCGCCACAAATGATGGGAAACAGCTGTGC 480  
 Qy 421 CCCCCGGGAAATCCAAAGTACCTTGAGAGAGATTAACAGAGATCTGGAACCCAAAGGGGG 540  
 Db 421 CCCCCGGGAAATCCAAAGTACCTTGAGAGAGATTAACAGAGATCTGGAACCCAAAGGGGG 540  
 Qy 541 AGCGACCTGAGGAGATGACGAGTAACCTCCCTCG 576  
 Db 541 AGCGACCTGAGGAGATGACGAGTAACCTCCCTCG 576

RESULT 2  
 ARI170503  
 LOCUS ARI170503 576 bp DNA linear PAT 17-DEC-2001  
 DEFINITION Sequence 1 from patent US 6291658.  
 ACCESSION ARI170503  
 VERSION ARI170503.1 GI:17908462  
 KEYWORDS  
 SOURCE Unknown.  
 ORGANISM Unknown.  
 UNCLASSIFIED.  
 REFERENCE 1 (bases 1 to 576)  
 AUTHORS Gure,A.O., Tureci,O., Sahin,U., Tsang,S., Scanlan,M.J., Knuth,A., Pfeundscher,M., Old,L.J. and Chen,Y.-T.  
 TITLE Isolated nucleic acid molecules encoding SSX family members and thereof  
 JOURNAL Patent: US 6291658-A 1 18-SEP-2001;  
 FEATURES Location/Qualifiers  
 source 1..576  
 /organism="unknown"  
 /mol\_type="unassigned DNA"

Query Match 100.0%; Score 576; DB 6; Length 576;  
 Best Local Similarity 100.0%; Pred. No. 2.6e-164;  
 Matches 576; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ATGAACGGAGACGCGCTTTGCAAGGAGACCCAGGAGTGTCTCAAAATATCAGAGAAG 60  
 Db 1 ATGAACGGAGACGCGCTTTGCAAGGAGACCCAGGAGTGTCTCAAAATATCAGAGAAG 60  
 Qy 61 TTACGAAAGGCGCTTCGATGATATGCGCAAAATCTCTTAAGAAAGAGTGGGAAAAGATG 120  
 Db 61 TTACGAAAGGCGCTTCGATGATATGCGCAAAATCTCTTAAGAAAGAGTGGGAAAAGATG 120  
 Qy 121 AAATCCTCGGAGAAAATCGTCTATGTATATGAGCTAAATCTATGAGGTCTATGACTAAA 180  
 Db 121 AAATCCTCGGAGAAAATCGTCTATGTATATGAGCTAAATCTATGAGGTCTATGACTAAA 180  
 Qy 181 CTAGGTTTCAGGTCACCGCTCCACCTTTTCATGCGTAGTAAACGGGCTGCAGACTTCCAC 240  
 Db 181 CTAGGTTTCAGGTCACCGCTCCACCTTTTCATGCGTAGTAAACGGGCTGCAGACTTCCAC 240  
 Qy 241 GGGAAATGATTTTGGTAACGATCGAAACCAACAGGAAATCAGGTTGAACGTCCTCAGATGACT 300  
 Db 241 GGGAAATGATTTTGGTAACGATCGAAACCAACAGGAAATCAGGTTGAACGTCCTCAGATGACT 300  
 Qy 301 TTGCGACGCTCCAGAGAAATCTTCCGAAAGATCATGCGCAAGAAATGATGGGAAACAGCTGTGC 420  
 Db 301 TTGCGACGCTCCAGAGAAATCTTCCGAAAGATCATGCGCAAGAAATGATGGGAAACAGCTGTGC 420  
 Qy 361 AATGTTTGAAGGAGTCCAGAGGATCTCGCCACAAATGATGGGAAACAGCTGTGC 480  
 Db 361 AATGTTTGAAGGAGTCCAGAGGATCTCGCCACAAATGATGGGAAACAGCTGTGC 480  
 Qy 421 CCCCCGGGAAATCCAAAGTACCTTGAGAGAGATTAACAGAGATCTGGAACCCAAAGGGGG 540  
 Db 421 CCCCCGGGAAATCCAAAGTACCTTGAGAGAGATTAACAGAGATCTGGAACCCAAAGGGGG 540  
 Qy 541 AGCGACCTGAGGAGATGACGAGTAACCTCCCTCG 576  
 Db 541 AGCGACCTGAGGAGATGACGAGTAACCTCCCTCG 576

RESULT 3  
 ARI182977  
 LOCUS ARI182977 576 bp DNA linear PAT 20-APR-2002  
 DEFINITION Sequence 1 from patent US 6339140.  
 ACCESSION ARI182977  
 VERSION ARI182977.1 GI:20226184  
 KEYWORDS  
 SOURCE Unknown.  
 ORGANISM Unknown.  
 UNCLASSIFIED.  
 REFERENCE 1 (bases 1 to 576)  
 AUTHORS Gure,A.O., Tureci,O., Sahin,U., Tsang,S., Scanlan,M.J., Knuth,A., Pfeundscher,M., Old,L.J. and Chen,Y.-T.  
 TITLE SSX family proteins  
 JOURNAL Patent: US 6339140-A 1 15-JAN-2002;  
 FEATURES Location/Qualifiers  
 source 1..576  
 /organism="unknown"  
 /mol\_type="unassigned DNA"

Query Match 100.0%; Score 576; DB 6; Length 576;  
 Best Local Similarity 100.0%; Pred. No. 2.6e-164;  
 Matches 576; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ATGAACGGAGACGCGCTTTGCAAGGAGACCCAGGAGTGTCTCAAAATATCAGAGAAG 60

Db 1 ATGACGGAGACGACGCCCTTTGCAAGGAGAGACCCAGGAGATGATGCTCAAATATCAGAGAAG 60  
QY 61 TTACGAAAGCCCTTCGATGATATTTGCCAAATATCTCTCTAAGAAAGAGTGGGAAAGATG 120  
Db 61 TTACGAAAGCCCTTCGATGATATTTGCCAAATATCTCTCTAAGAAAGAGTGGGAAAGATG 120  
QY 121 AAATCCTCGAGAAAATCGTCTATGTTATATGAAGCTAAACTATGAGGTCATGACTTAA 180  
Db 121 AAATCCTCGAGAAAATCGTCTATGTTATATGAAGCTAAACTATGAGGTCATGACTTAA 180  
QY 181 CTAGGTTTCAAGTCAACCTCCACCTTTCATGCGTAGTAAACGGGCTGCAGACTTCCAC 240  
Db 181 CTAGGTTTCAAGTCAACCTCCACCTTTCATGCGTAGTAAACGGGCTGCAGACTTCCAC 240  
QY 241 GGAATGATTTTGGTAACGATCGAAACACAGGAATCAGGTTTGAACCTCTCAGATGACT 300  
Db 241 GGAATGATTTTGGTAACGATCGAAACACAGGAATCAGGTTTGAACCTCTCAGATGACT 300  
QY 301 TTCCGAGCCTCCAGAGAAATCTTCCCGAAGATCATGCCAAGAACCCAGAGAGGAAGAA 360  
Db 301 TTCCGAGCCTCCAGAGAAATCTTCCCGAAGATCATGCCAAGAACCCAGAGAGGAAGAA 360  
QY 361 AATGGTTTGAAGAGTGCAGAGGATCTGGCCACAAATGATGGAAACAGCTGTGC 420  
Db 361 AATGGTTTGAAGAGTGCAGAGGATCTGGCCACAAATGATGGAAACAGCTGTGC 420  
QY 421 CCCCCGGGAAATCCAAAGTACCTTGGAGAAGATTAAACAAGACATCTGGACCCCAAGGGGG 480  
Db 421 CCCCCGGGAAATCCAAAGTACCTTGGAGAAGATTAAACAAGACATCTGGACCCCAAGGGGG 480  
QY 481 AAACATGCTGGACCCACAGACTGCGTGAGAGAAAGCAGCTGGTGGTTTATGAAGATC 540  
Db 481 AAACATGCTGGACCCACAGACTGCGTGAGAGAAAGCAGCTGGTGGTTTATGAAGATC 540  
QY 541 AGCGACCTGAGGAGATGACGAGTAACCTCCCTCG 576  
Db 541 AGCGACCTGAGGAGATGACGAGTAACCTCCCTCG 576

RESULT 4  
AR305671  
LOCUS AR305671 576 bp DNA linear PAT 12-JUN-2003  
DEFINITION Sequence 5 from patent US 6548064.  
ACCESSION AR305671  
VERSION AR305671.1 GI:31695155  
KEYWORDS Unknown.  
SOURCE Unknown.  
ORGANISM Unclassified.  
REFERENCE 1 (bases 1 to 576)  
AUTHORS Tureci, O., Sahin, U., Pfreundschuh, M., Rammensee, H.G. and Stevanovic, S.  
TITLE Isolated peptides consisting of amino acid sequences found in SSX or NY-ESO-1 molecules, which bind to HLA molecules  
JOURNAL Patent: US 6548064-A 5 15-APR-2003;  
FEATURES  
source Location/Qualifiers  
1..576  
/organism="unknown"  
/mol\_type="genomic DNA"  
ORIGIN  
Query Match 100.0%; Score 576; DB 6; Length 576;  
Best Local Similarity 100.0%; Pred. No. 2.6e-164;  
Matches 576; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 ATGACGGAGACGACGCCCTTTGCAAGGAGAGACCCAGGAGATGATGCTCAAATATCAGAGAAG 60  
Db 1 ATGACGGAGACGACGCCCTTTGCAAGGAGAGACCCAGGAGATGATGCTCAAATATCAGAGAAG 60  
QY 61 TTACGAAAGCCCTTCGATGATATTTGCCAAATATCTCTCTAAGAAAGAGTGGGAAAGATG 120  
Db 61 TTACGAAAGCCCTTCGATGATATTTGCCAAATATCTCTCTAAGAAAGAGTGGGAAAGATG 120

QY 121 AAATCCTCGAGAAAATCGTCTATGTTATATGAAGCTAAACTATGAGGTCATGACTTAA 180  
Db 121 AAATCCTCGAGAAAATCGTCTATGTTATATGAAGCTAAACTATGAGGTCATGACTTAA 180  
QY 181 CTAGGTTTCAAGTCAACCTCCACCTTTCATGCGTAGTAAACGGGCTGCAGACTTCCAC 240  
Db 181 CTAGGTTTCAAGTCAACCTCCACCTTTCATGCGTAGTAAACGGGCTGCAGACTTCCAC 240  
QY 241 GGAATGATTTTGGTAACGATCGAAACACAGGAATCAGGTTTGAACCTCTCAGATGACT 300  
Db 241 GGAATGATTTTGGTAACGATCGAAACACAGGAATCAGGTTTGAACCTCTCAGATGACT 300  
QY 301 TTCCGAGCCTCCAGAGAAATCTTCCCGAAGATCATGCCAAGAACCCAGAGAGGAAGAA 360  
Db 301 TTCCGAGCCTCCAGAGAAATCTTCCCGAAGATCATGCCAAGAACCCAGAGAGGAAGAA 360  
QY 361 AATGGTTTGAAGAGTGCAGAGGATCTGGCCACAAATGATGGAAACAGCTGTGC 420  
Db 361 AATGGTTTGAAGAGTGCAGAGGATCTGGCCACAAATGATGGAAACAGCTGTGC 420  
QY 421 CCCCCGGGAAATCCAAAGTACCTTGGAGAAGATTAAACAAGACATCTGGACCCCAAGGGGG 480  
Db 421 CCCCCGGGAAATCCAAAGTACCTTGGAGAAGATTAAACAAGACATCTGGACCCCAAGGGGG 480  
QY 481 AAACATGCTGGACCCACAGACTGCGTGAGAGAAAGCAGCTGGTGGTTTATGAAGATC 540  
Db 481 AAACATGCTGGACCCACAGACTGCGTGAGAGAAAGCAGCTGGTGGTTTATGAAGATC 540  
QY 541 AGCGACCTGAGGAGATGACGAGTAACCTCCCTCG 576  
Db 541 AGCGACCTGAGGAGATGACGAGTAACCTCCCTCG 576

## RESULT 5

AR305671  
LOCUS AR305671 576 bp mRNA linear PAT 08-MAY-2003  
DEFINITION Sequence 1 from Patent EP1300463.  
ACCESSION AR305671  
VERSION AR305671.1 GI:30519358  
KEYWORDS Homo sapiens (human)  
SOURCE Homo sapiens  
ORGANISM Homo sapiens  
REFERENCE 1  
AUTHORS Gure, A.O., Tureci, O., Sahin, U., Tsang, S., Scanlan, M.J., Knuth, A., Pfreundschuh, M., Old, L.J. and Chen, Y.T.  
TITLE Isolated nucleic acid molecules encoding SSX family members and uses thereof  
JOURNAL Patent: EP 1300463-A 1 09-APR-2003;  
LUDWIG INSTITUTE FOR CANCER RESEARCH (US); MEMORIAL SLOAN-KETTERING CANCER CENTER (US); CORNELL RESEARCH FOUNDATION (US)

FEATURES  
source Location/Qualifiers  
1..576  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
gene 1..576  
/gene="SSX4"  
CDS 1..567  
/genes="SSX4"  
/codon\_start=1  
/protein\_id="CAD90570.1"  
/db\_xref="GI:30519359"  
/translation="MNGDDAFARPRDDAQISEKLKAFEDDIKAFESKEWEKMKSE  
KIVYVYMLNYEYVTKLGFVTLPPPMRSKRAADFGNDNRHNRNQRVQMTFG  
SLQRIPEKTKPKPEEENGLKEVPEASGPNQDKQLCPPGNPSTLEKINTKSGPKRG  
KXAWHLRLERKQLVYVEISDPEEDDE"  
ORIGIN  
Query Match 100.0%; Score 576; DB 6; Length 576;

Best Local Similarity 100.0%; Pred. No. 2.6e-164;  
Matches 576; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGACGAGAGCGCCCTTTGCAAGGAGAGCCAGGATGATGCTCAATATCAGAGAAG 60  
DB 1 ATGACGAGAGCGCCCTTTGCAAGGAGAGCCAGGATGATGCTCAATATCAGAGAAG 60

QY 61 TTACGAAAGGCGCTTCGATGATATGCGCAATATCTCTTAAGAAAGAGTGGAAAGATG 120  
DB 61 TTACGAAAGGCGCTTCGATGATATGCGCAATATCTCTTAAGAAAGAGTGGAAAGATG 120

QY 121 AAATCTCGGAGAAATGCTCTATGTATATGAAGTAACTATGAGTCAATGACTAAA 180  
DB 121 AAATCTCGGAGAAATGCTCTATGTATATGAAGTAACTATGAGTCAATGACTAAA 180

QY 181 CTAGGTTTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAG 240  
DB 181 CTAGGTTTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAG 240

QY 241 GCGAATGATTTTGGTAAACGATCGAAACACAGGAATCAGGTTGAACGCTTCAGATGACT 300  
DB 241 GCGAATGATTTTGGTAAACGATCGAAACACAGGAATCAGGTTGAACGCTTCAGATGACT 300

QY 301 TTCGACAGGCTTCAGAGATCTTCCGAGATCATGCCCCAGAGAGCCAGAGAGAGAA 360  
DB 301 TTCGACAGGCTTCAGAGATCTTCCGAGATCATGCCCCAGAGAGCCAGAGAGAGAA 360

QY 361 AATGTTTGAAGGAGTGCAGAGGATCTGCGCCACAAATGATGGGAAACAGCTGTGC 420  
DB 361 AATGTTTGAAGGAGTGCAGAGGATCTGCGCCACAAATGATGGGAAACAGCTGTGC 420

QY 421 CCCCAGGAAATCCAAAGTACCTTGGAGAGATTAACAGACATCTGGACCCAAAGGGGG 480  
DB 421 CCCCAGGAAATCCAAAGTACCTTGGAGAGATTAACAGACATCTGGACCCAAAGGGGG 480

QY 481 AAACATGCTCGACCCAGAGATCTGCGCCACAAATGATGGGAAACAGCTGTGC 540  
DB 481 AAACATGCTCGACCCAGAGATCTGCGCCACAAATGATGGGAAACAGCTGTGC 540

QY 541 AGCGACCTCGAGAGAGATGACGAGTAATCCCTCG 576  
DB 541 AGCGACCTCGAGAGAGATGACGAGTAATCCCTCG 576

BD096872 576 bp DNA linear PAT 27-AUG-2002  
Isolated nucleic acid molecules encoding SSX family members and  
uses thereof.

BD096872  
VERSION BD096872.1 GI:226242460  
KEYWORDS JP 2001527408-A/1.  
SOURCE unidentified  
ORGANISM unidentified

REFERENCE  
AUTHORS Gure,A.O., Tureci,O., Sahin,U., Tsang,S., Scanlan,M.J., Knuth,A., Pfreundschuh,M., Old,L.J. and Chen,Y.T.  
TITLE Isolated nucleic acid molecules encoding SSX family members and uses thereof

JOURNAL Patent: JP 2001527408-A 1 25-DEC-2001;  
LUDWIG INSTITUTE FOR CANCER RESEARCH, MEMORIAL SLOAN KETTERING  
CANCER CENTER, CORNELL RESEARCH FOUNDATION  
OS Unidentified  
PN JP 2001527408-A/1  
PD 25-DEC-2001  
PP 25-FEB-1998 JP 1998548050  
PR 05-MAY-1997 US 08/851138  
PI ALI O GURE, OZLEM TURECI, UGUR SAHIN, SOLAM TSANG, MATTHEW J PI  
SCANLAN,  
PI ALEXANDER KNUTH, MICHAEL PFEUNDSCUH, LLOYD J OLD, YAO TSENG PI  
CHEN  
PC C12N5/10, C12N15/12, C12P21/02, C12Q1/68

CC Strandedness: Single;  
CC Topology: Linear;  
CC Isolated nucleic acid molecules encoding SSX family members  
CC thereof  
FH Key Location/Qualifiers  
FT source 1..576  
/organism="unidentified"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:32644"

FEATURES  
source  
1..576  
/organism="unidentified"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:32644"

ORIGIN  
Query Match 100.0%; Score 576; DB 6; Length 576;  
Best Local Similarity 100.0%; Pred. No. 2.6e-164;  
Matches 576; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGACGAGAGCGCCCTTTGCAAGGAGAGCCAGGATGATGCTCAATATCAGAGAAG 60  
DB 1 ATGACGAGAGCGCCCTTTGCAAGGAGAGCCAGGATGATGCTCAATATCAGAGAAG 60

QY 61 TTACGAAAGGCGCTTCGATGATATGCGCAATATCTCTTAAGAAAGAGTGGAAAGATG 120  
DB 61 TTACGAAAGGCGCTTCGATGATATGCGCAATATCTCTTAAGAAAGAGTGGAAAGATG 120

QY 121 AAATCTCGGAGAAATGCTCTATGTATATGAAGTAACTATGAGTCAATGACTAAA 180  
DB 121 AAATCTCGGAGAAATGCTCTATGTATATGAAGTAACTATGAGTCAATGACTAAA 180

QY 181 CTAGGTTTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAG 240  
DB 181 CTAGGTTTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAG 240

QY 241 GCGAATGATTTTGGTAAACGATCGAAACACAGGAATCAGGTTGAACGCTTCAGATGACT 300  
DB 241 GCGAATGATTTTGGTAAACGATCGAAACACAGGAATCAGGTTGAACGCTTCAGATGACT 300

QY 301 TTCGACAGGCTTCAGAGATCTTCCGAGATCATGCCCCAGAGAGCCAGAGAGAGAA 360  
DB 301 TTCGACAGGCTTCAGAGATCTTCCGAGATCATGCCCCAGAGAGCCAGAGAGAGAA 360

QY 361 AATGTTTGAAGGAGTGCAGAGGATCTGCGCCACAAATGATGGGAAACAGCTGTGC 420  
DB 361 AATGTTTGAAGGAGTGCAGAGGATCTGCGCCACAAATGATGGGAAACAGCTGTGC 420

QY 421 CCCCAGGAAATCCAAAGTACCTTGGAGAGATTAACAGACATCTGGACCCAAAGGGGG 480  
DB 421 CCCCAGGAAATCCAAAGTACCTTGGAGAGATTAACAGACATCTGGACCCAAAGGGGG 480

QY 481 AAACATGCTCGACCCAGAGATCTGCGCCACAAATGATGGGAAACAGCTGTGC 540  
DB 481 AAACATGCTCGACCCAGAGATCTGCGCCACAAATGATGGGAAACAGCTGTGC 540

QY 541 AGCGACCTCGAGAGAGATGACGAGTAATCCCTCG 576  
DB 541 AGCGACCTCGAGAGAGATGACGAGTAATCCCTCG 576

RESULT 7  
BD218420  
LOCUS  
DEFINITION  
SSX gene, method of determining the occurrence of cancer in sample  
by determining the expression of peptides originating in the SSX  
gene and NY-ESO-1 gene and utilization thereof.

ACCESSION  
BD218420  
VERSION  
JP 2002519013-A/5.  
KEYWORDS  
Homo sapiens (human)  
SOURCE  
ORGANISM  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.



```
REFERENCE
AUTHORS Tureci,O., Sahin,U., Pfreundschuh,M., Rammensee,G., Stevanovic,S.,
        Chen,Y.T., Gure,A. and Old,L.J.
TITLE   SSX gene, method of determining the occurrence of cancer in sample
        by determining the expression of peptides originating in the SSX
        gene and NY-ESO-1 gene and utilization thereof
JOURNAL LUDWIG INSTITUTE FOR CANCER RESEARCH
COMMENT  OS Homo sapiens (human)
        PN JP 2002519013-A/5
        PD 02-JUL-2002
        PF 25-JUN-1999 JP 2000557145
        PR 26-JUN-1998 US 09/105839
        PI OZLEM TURECI,UGUR SAHIN,MICHAEL PFEUNDSCHUH,GEORG RAMMENSEE,
        STEFAN STEVANOVIC,YAO TSENG CHEN,ALI GURE,LLOYD J OLD PC
        C12N15/09,A61K38/00,C07H21/04,C12N5/10,C12P21/04,C12Q1/58, PC
        G01N33/48,
        PC G01N33/53,G01N33/574,C12N15/00,A61K37/02,C12N5/00 CC SSX
        gene, method of determining the occurrence of cancer in CC
        sample by
        CC determining the expression of peptides originating in the SSX
        gene and
        CC NY-ESO-1 gene and utilization thereof
        FH Key
        FT source
        FT Location/Qualifiers
        /organism='Homo sapiens (human)'.
        1..576
        /organism='Homo sapiens'
        /mol_type='genomic DNA'
        /db_xref='taxon:9606'

ORIGIN
Query Match 100.0%; Score 576; DB 6; Length 576;
Best Local Similarity 100.0%; Pred. No. 2.6e-164;
Matches 576; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGAACCGAGACGACGCTTTGCAAGGAGACCCAGGATGATGCTCAAAATATCAGAGAAG 60
DB 1 ATGAACCGAGACGACGCTTTGCAAGGAGACCCAGGATGATGCTCAAAATATCAGAGAAG 60
QY 61 TTACGAAAGGCTTCGATGATATTTGCCAAATATCTTCTCTAAGAAAGAGTGGGAAAAGATG 120
DB 61 TTACGAAAGGCTTCGATGATATTTGCCAAATATCTTCTCTAAGAAAGAGTGGGAAAAGATG 120
QY 121 AAATCCTCGGAGAAATCGTCTATGCTATATGAAGCTAACTATGAGGTCATGACTAAA 180
DB 121 AAATCCTCGGAGAAATCGTCTATGCTATATGAAGCTAACTATGAGGTCATGACTAAA 180
QY 181 CTAGGTTTCAAGGTCACCCCTCCCACTTTCATGCTAGTAAACGGGCTGCAGACTTCCAC 240
DB 181 CTAGGTTTCAAGGTCACCCCTCCCACTTTCATGCTAGTAAACGGGCTGCAGACTTCCAC 240
QY 241 GGGATGATTTGGTACGATCGAAACACACGAGATCAGGTTGAACTCCTCAGATGACT 300
DB 241 GGGATGATTTGGTACGATCGAAACACACGAGATCAGGTTGAACTCCTCAGATGACT 300
QY 301 TTGCGGACGCTCCAGAGAACTTCCCGAAGATCATGCCAGAGATCCAGAGCCAGAGAGAA 360
DB 301 TTGCGGACGCTCCAGAGAACTTCCCGAAGATCATGCCAGAGATCCAGAGCCAGAGAGAA 360
QY 361 AATGGTTTGAAGGAAGTGCACAGAGGCTCTGGCCCAACAAATATGATGGGAAACAGCTCTGC 420
DB 361 AATGGTTTGAAGGAAGTGCACAGAGGCTCTGGCCCAACAAATATGATGGGAAACAGCTCTGC 420
QY 421 CCCCCGGGAATCCAACTACCTTGGAGAAGATTAAAGACATCTGGAACCCAAAGGGGG 480
DB 421 CCCCCGGGAATCCAACTACCTTGGAGAAGATTAAAGACATCTGGAACCCAAAGGGGG 480
QY 481 AAACATGCTGGACCCACAGCTCGGTGAGAGAAAGCAGCTGGTGGTTTATGAGAGATC 540
DB 481 AAACATGCTGGACCCACAGCTCGGTGAGAGAAAGCAGCTGGTGGTTTATGAGAGATC 540

103 541 AGCGACCTCAGGAGATGACGAGTACTCCCTCCG 576
104 541 AGCGACCTCAGGAGATGACGAGTACTCCCTCCG 576

RESULT 8
LOCUS HSU90841 576 bp mRNA linear PRI 18-MAR-1998
DEFINITION Homo sapiens SSX4 (SSX4) mRNA, complete cds.
ACCESSION U90841
VERSION U90841.1 GI:2952022
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 576)
Gure,A.O., Tureci,O., Sahin,U., Tsang,S., Scanlan,M.J., Jager,E.,
Knuth,A., Pfreundschuh,M., Old,L.J. and Chen,Y.T.
SSX: a multigene family with several members transcribed in normal
testis and human cancer
Int. J. Cancer 72 (6), 965-971 (1997)
JOURNAL 98021352
MEDLINE 9378559
PUBMED
REFERENCE 2 (bases 1 to 576)
Gure,A.O., Tureci,O., Sahin,U., Tsang,S., Scanlan,M., Knuth,A.,
Pfreundschuh,M., Old,L.J. and Chen,Y.T.
Direct Submission
Submitted (05-FEB-1997) Pathology C-320, Cornell University, 1300
York Ave., New York City, NY 10021, USA
Location/Qualifiers
1..576
/organism='Homo sapiens'
/mol_type='mRNA'
/db_xref='taxon:9606'
1..576
/genes='SSX4'
1..567
/genes='SSX4'
/notes='putative translation product; Kruppel-associated
box containing SSX gene'
/codon_start=1
/product='SSX4'
/protein_id='AAC05820.1'
/db_xref='GI:2952023'
/translation='MNGDDAFARPRDDAQISEKLKAFDDIAKYPSKKEWKMSSE
KIVIVYIMKLNIYVMTKLGFKVTLPPFMRSGRAADPHGNDGNDNRHNRKQVERQMTFG
SQRIPIPKMPKPEEENGLKEVPBASGQNDGKQLCPGPNSTLEKINKTSGPKRG
KHAWTHRLRERKQLVYVERISDPEEDE'
1..18
/genes='SSX4'
/notes='corresponds to the 5' PCR primer that was used to
isolate this gene and is identical to SSX2, GenBank
Accession Number X86175; therefore, the SSX4 gene sequence
might contain differences within this span'

primer_bind
Query Match 100.0%; Score 576; DB 9; Length 576;
Best Local Similarity 100.0%; Pred. No. 2.6e-164;
Matches 576; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGAACCGAGACGACGCTTTGCAAGGAGACCCAGGATGATGCTCAAAATATCAGAGAAG 60
DB 1 ATGAACCGAGACGACGCTTTGCAAGGAGACCCAGGATGATGCTCAAAATATCAGAGAAG 60
QY 61 TTACGAAAGGCTTCGATGATATTTGCCAAATATCTTCTCTAAGAAAGAGTGGGAAAAGATG 120
DB 61 TTACGAAAGGCTTCGATGATATTTGCCAAATATCTTCTCTAAGAAAGAGTGGGAAAAGATG 120
QY 121 AAATCCTCGGAGAAATCGTCTATGCTATGCTATGAGCTAACTATGAGGTCATGACTAAA 180
DB 121 AAATCCTCGGAGAAATCGTCTATGCTATGCTATGAGCTAACTATGAGGTCATGACTAAA 180
```

181 CTAGGTTTCAAGGTCACCCCTCCACCTTTCATGGGTAGTAAACGGGCTCAGACTTCCAC 240  
 Db CTAGGTTTCAAGGTCACCCCTCCACCTTTCATGGGTAGTAAACGGGCTCAGACTTCCAC 240  
 241 GGGAAATGATTTGGTAAAGATCGAAACACAGGAATCAGGTTGAAGCTCTCAGACTACT 300  
 Db GGGAAATGATTTGGTAAAGATCGAAACACAGGAATCAGGTTGAAGCTCTCAGACTACT 300  
 301 TTCGGCAGCCCTCCAGAGAAATCTTCCGGAAGATCATGCCAAGAAGCCAGCAGAGGAAGAA 360  
 Db TTCGGCAGCCCTCCAGAGAAATCTTCCGGAAGATCATGCCAAGAAGCCAGCAGAGGAAGAA 360  
 361 AATGTTTGAAGAGAGTCCAGAGCATCTGCCCCACAAATATGATGGAAACACTGTGC 420  
 Db AATGTTTGAAGAGAGTCCAGAGCATCTGCCCCACAAATATGATGGAAACACTGTGC 420  
 421 CCCCCGGGAATCAAGTACCTTGGAGAGATTACAGACATCTGACCCCAAAAGGGGG 480  
 Db CCCCCGGGAATCAAGTACCTTGGAGAGATTACAGACATCTGACCCCAAAAGGGGG 480  
 481 AAACATGCTGACCCACAGAGTGGTGCAGAGAAAGAGCTGGTGTATGAAGAGATC 540  
 Db AAACATGCTGACCCACAGAGTGGTGCAGAGAAAGAGCTGGTGTATGAAGAGATC 540  
 541 AGCAGCCTGAGGAGATGACAGTAATCTCCCTCG 576  
 Db AGCAGCCTGAGGAGATGACAGTAATCTCCCTCG 576

RESULT 9  
 BC005325  
 LOCUS  
 DEFINITION Homo sapiens synovial sarcoma, X breakpoint 4, transcript variant 1, mRNA (cdna clone MGC:12411 IMAGE:3961688), complete cds.  
 ACCSSION BC005325  
 VERSION BC005325.1 GI:13529094  
 KEYWORDS MGC.  
 SOURCE Homo sapiens  
 ORGANISM Homo sapiens (human)  
 Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 1 (bases 1 to 1250)  
 Strausberg,R.L., Feingold,E.A., Grouse,L.H., Derge,J.G., Klausner,R.D., Collins,F.S., Wagner,L., Shenmen,C.M., Schuler,G.D., Altschul,S.F., Zeeberg,B., Buetow,K.H., Schaefer,C.F., Bhat,N.K., Hopkins,R.F., Jordan,H., Moore,T., Max,S.I., Wang,J., Hsieh,F., Diatchenko,L., Marusina,K., Farmer,A.A., Rubin,G.M., Hong,L., Stapleton,M., Soares,M.B., Bonaldo,M.F., Casavant,T.L., Scherz,E., Brownstein,M.J., Usdin,T.B., Ioshizuka,S., Carninci,P., Frange,C., Raha,S., Loquellano,N.A., Peters,G.J., Abramson,K.J., Mullahy,S.J., Bosak,S.A., McEwan,P.J., McKernan,K.J., Malek,J.A., Gunaratne,P.H., Richards,S., Villalón,D.K., Muzny,D.M., Sodergren,E.J., Lu,X., Gibbs,R.A., Fahey,J., Helton,E., Kettman,M., Madan,A., Rodriguez,S., Sanchez,A., Whiting,M., Madan,A., Young,A.C., Shevchenko,Y., Bouffard,G.G., Blakesley,R.W., Touchman,J.W., Green,E.D., Dickson,M.C., Rodriguez,A.C., Grimwood,J., Schmutz,J., Myers,R.M., Butlerfield,Y.S., Krzyzanski,M.I., Skalska,U., Smalhus,D.E., Schnerch,A., Schein,J.E., Jones,S.J. and Marra,M.A.  
 Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences  
 Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)  
 22388257  
 12477932  
 2 (bases 1 to 1250)  
 Strausberg,R.  
 Direct Submission  
 Submitted (27-MAR-2001) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA  
 NIH-MGC Project URL: <http://mgc.nci.nih.gov>

COMMENT  
 Contact: MGC help desk  
 Email: [gcaps-k@mail.nih.gov](mailto:gcaps-k@mail.nih.gov)  
 Tissue Procurement: ATCC  
 cDNA Library Preparation: CLONTECH Laboratories, Inc.  
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (ILNL)  
 DNA Sequencing by: Sequencing Group at the Stanford Human Genome Center, Stanford University School of Medicine, Stanford, CA 94305  
 Web site: <http://www-shgc.stanford.edu>  
 Contact: (Dickson, Mark) [mcd@paxil.stanford.edu](mailto:mcd@paxil.stanford.edu)  
 Dickson, M., Schmutz, J., Grimwood, J., Rodriguez, A., and Myers, R. M.  
 Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/ILNL at: <http://image.llnl.gov>  
 Series: IRAL Plate: 16 Row: m Column: 20  
 This clone was selected for full length sequencing because it passed the following selection criteria: matched mRNA gi: 28559011.  
 location/Qualifiers  
 1. .1250  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="MGC:12411 IMAGE:3961688"  
 /tissue\_type="Bladder, carcinoma"  
 /clone\_lib="NIH\_MGC\_53"  
 /lab\_host="DH10B"  
 /note="Vector: pDNR-LIB"  
 1. .1250  
 /gene="SSX4"  
 /note="synonym: MGC12411"  
 /db\_xref="LocusID:6759"  
 /db\_xref="MIM:300326"  
 59. .625  
 /codon\_start=1  
 /product="synovial sarcoma, X breakpoint 4, isoform a"  
 /protein\_id="AAH05325.1"  
 /db\_xref="GI:13529095"  
 /translation="MNGDDAFARRPRDDAQISEKLRKAFDDIAYFKSEWEMKMSSEKIVVYMKLVNVEVMTKLGPVTLPPPMRSKRAADFGNDGNDRNHRNQVERPOMTFGLQRIEPIKMPKPEASEENGLKEVPEASGPNQDKQLCPNPSTLEKINKTSPEKRGKHWTHRLRERKQLVVEEISDFEEDDE"  
 125. .304  
 /note="KRAB; Region: krueppel associated box"  
 /db\_xref="CDD:smart00349"  
 misc\_feature  
 125. .304  
 /note="KRAB; Region: krueppel associated box"  
 /db\_xref="CDD:smart00349"  
 ORIGIN  
 Query Match 99.7%; Score 574.4; DB 9; Length 1250;  
 Best Local Similarity 99.8%; Pred. No. 8.8e-164;  
 Matches 575; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
 1 ATGAACGAGACGACGCGCTTTGCAAGGAGAGACCCAGGATGATGCTCAATATCAGAGAAG 60  
 Db 59 ATGAACGAGACGACGCGCTTTGCAAGGAGAGACCCAGGATGATGCTCAATATCAGAGAAG 118  
 61 TTACGAAAGCGCTTCGATGATATTTGCCAAATCTCTCTTAAGAAAGAGTGGGAAAGATG 120  
 Db 119 TTACGAAAGCGCTTCGATGATATTTGCCAAATCTCTCTTAAGAAAGAGTGGGAAAGATG 178  
 121 AAATCTCTCGGAGAAATCGTCTATGTATGATGAAGCTAAACTATGAGGTCATGACTAAA 180  
 Db 179 AAATCTCTCGGAGAAATCGTCTATGTATGATGAAGCTAAACTATGAGGTCATGACTAAA 238  
 181 CTAGGTTTCAAGGTCACCCCTCCACCTTTCATGGGTAGTAAACGGGCTCAGACTTCCAC 240  
 Db 239 CTAGGTTTCAAGGTCACCCCTCCACCTTTCATGGGTAGTAAACGGGCTCAGACTTCCAC 298  
 241 GGGAAATGATTTGGTAAAGATCGAAACACAGGAATCAGGTTGAAGCTCTCAGACTACT 300  
 Db 299 GGGAAATGATTTGGTAAAGATCGAAACACAGGAATCAGGTTGAAGCTCTCAGACTACT 358  
 301 TTCCGCGAGCTCCAGAGAAATCTTCCCGAAGATCATGCCAAGAAGCCAGCAGAGGAAGAA 360

Db 359 TTCCGACGCTCCAGAGATCTTCCCGAAGATCATGCCAAGACCCAGCAGAGAGAA 418  
QY 361 AATGGTTTGAAGGAAGTCCAGAGGCATCTGGCCCAACAATATGGAACACCTGTGC 420  
Db 419 AATGGTTTGAAGGAAGTCCAGAGGCATCTGGCCCAACAATATGGAACACCTGTGC 478  
QY 421 CCCCCGGGAATCCAAGTACCTTGGAGAAGATTAAAGACATCTGGACCCCAAAAGGGG 480  
Db 479 CCCCCGGGAATCCAAGTACCTTGGAGAAGATTAAAGACATCTGGACCCCAAAAGGGG 538  
QY 481 AAACATGCTGGACCCCAAGACATCTGGAGAAGATTAAAGACATCTGGAGAGATC 540  
Db 539 AAACATGCTGGACCCCAAGACATCTGGAGAAGATTAAAGACATCTGGAGAGATC 598  
QY 541 AGCACCCTGAGGAAGATGACGAGTAACCTCCCTCG 576  
Db 599 AGCACCCTGAGGAAGATGACGAGTAACCTCCCTCG 634

RESULT 10  
LOCUS A48452 766 bp DNA linear PAT 07-MAR-1997  
DEFINITION Sequence 17 from Patent WO9602641.  
ACCESSION A48452  
VERSION A48452.1 GI:2302236  
KEYWORDS unidentified  
SOURCE unclassified.  
ORGANISM unclassified.  
REFERENCE 1 (bases 1 to 766)  
AUTHORS Cooper, C.S. and Gusterson, B.A.  
TITLE MATERIALS AND METHODS RELATING TO THE DIAGNOSIS AND PROPHYLACTIC  
JOURNAL AND THERAPEUTIC TREATMENT OF SYNOVIAL SARCOMA  
PATENT: WO 9602641-A 17 01-FEB-1996;  
CANCER RES CAMPAIGN TECH (GB)  
COMMENT Other publication AU 2986595 960216.  
FEATURES  
Location/Qualifiers  
source 1..766  
/organism="unidentified"  
/mol\_type="unassigned DNA"  
/db\_xref="taxon:32644"

ORIGIN  
Query Match 86.2%; Score 496.6; DB 6; Length 766;  
Best Local Similarity 91.5%; Pred. No. 4.6e-140;  
Matches 526; Conservative 0; Mismatches 49; Indels 0; Gaps 0;  
QY 1 ATGAACGGAGACGACGCTTTGCAAGAGAGACCCAGGATGATGCTCAAAATACAGAAAG 60  
Db 92 ATGAACGGAGACGACGCTTTGCAAGAGAGACCCAGGATGATGCTCAAAATACAGAAAG 151  
QY 61 TTACGAAAGGCTTCGATGATTTGCCAAATATCTCTAAGAAAGAGTGGAAAAGATG 120  
Db 152 ATCCAAAAGGCTTCGATGATTTGCCAAATATCTCTAAGAAAGAGTGGAAAAGATG 211  
QY 121 AAATCCTCGAGAAAATCGTCTATGTATATGAAGCTAAATATGAGGTCTAGCTATAA 180  
Db 212 AAAGCCTCGAGAAAATCTCTATGTATATGAAGAAAGATGATGAGGTCTAGCTATAA 271  
QY 181 CTAGGTTTCAAGTCAACCTCCACCTTTCATGCTAGTAAACGGGTGAGACTTCCAC 240  
Db 212 AAAGCCTCGAGAAAATCTCTATGTATATGAAGAAAGATGATGAGGTCTAGCTATAA 271  
QY 181 CTAGGTTTCAAGTCAACCTCCACCTTTCATGCTAGTAAACGGGTGAGACTTCCAC 240  
Db 272 CTAGGTTTCAAGGCAACCTCCACCTTTCATGCTAGTAAACGGGTGAGACTTCCAC 331  
QY 241 GGGAAATGATTTGGTAACGATCGAAACCAAGGATCAGGTTGAACGCTCTCAGATGACT 300  
Db 272 CTAGGTTTCAAGGCAACCTCCACCTTTCATGCTAGTAAACGGGTGAGACTTCCAC 331  
QY 241 GGGAAATGATTTGGTAACGATCGAAACCAAGGATCAGGTTGAACGCTCTCAGATGACT 300  
Db 332 GGGAAATGATTTGGTAACGATCGAAACCAAGGATCAGGTTGAACGCTCTCAGATGACT 391  
QY 301 TTCCGGACGCTCCAGAGAAATCTTCCCGAAGATCATGCCAAGAGCCAGCAGAGGAAGAA 360  
Db 392 GGGAAATGATTTGGTAACGATCGAAACCAAGGATCAGGTTGAACGCTCTCAGATGACT 391  
QY 301 TTCCGGACGCTCCAGAGAAATCTTCCCGAAGATCATGCCAAGAGCCAGCAGAGGAAGAA 360  
Db 392 TTCCGGACGCTCCAGAGAAATCTTCCCGAAGATCATGCCAAGAGCCAGCAGAGGAAGAA 451  
QY 361 AATGGTTTGAAGGAAGTCCAGAGGCATCTGGCCCAACAATATGGAACACCTGTGC 420

Db 452 AATGATTCGAGCAAGTGCAGAAAGCATCTGGCCCAACAATATGATGGAAGAGACTGTGC 511  
QY 421 CCCCCGGGAATCCAAGTACCTTGGAGAAGATTAAAGACATCTGGACCCCAAAAGGGG 480  
Db 512 CCCCCGGGAACCAACTCTCTGAGAAGATTACAGAGATCTGGACCCCAAAAGGGG 571  
QY 481 AAACATGCTGGACCCCAAGACATCTGGAGAAGATTAAAGACATCTGGAGAGATC 540  
Db 572 GAACATGCTGGACCCCAAGACATCTGGAGAAGATTAAAGACATCTGGAGAGATC 631  
QY 541 AGCACCCTGAGGAAGATGACGAGTAACCTCCCTCG 575  
Db 632 AGCACCCTGAGGAAGATGACGAGTAACCTCCCTCG 666

RESULT 11  
LOCUS AX821947 766 bp DNA linear PAT 10-DEC-2003  
DEFINITION Sequence 75 from Patent WO03068961.  
ACCESSION AX821947  
VERSION AX821947.1 GI:39725168  
KEYWORDS Homo sapiens (human)  
SOURCE Homo sapiens  
ORGANISM Homo sapiens  
REFERENCE 1  
AUTHORS Andrews, P.A., Walsh, J.A. and Gokhale, P.A.  
TITLE Method to modify differentiation of pluripotential stem cells  
JOURNAL Patent: WO 03068961-A 75 21-AUG-2003;  
Axordia Limited (GB)  
FEATURES  
Location/Qualifiers  
source 1..766  
/organism="Homo sapiens"  
/mol\_type="unassigned DNA"  
/db\_xref="taxon:9606"

ORIGIN  
Query Match 86.2%; Score 496.6; DB 6; Length 766;  
Best Local Similarity 91.5%; Pred. No. 4.6e-140;  
Matches 526; Conservative 0; Mismatches 49; Indels 0; Gaps 0;  
QY 1 ATGAACGGAGACGACGCTTTGCAAGAGAGACCCAGGATGATGCTCAAAATACAGAAAG 60  
Db 92 ATGAACGGAGACGACGCTTTGCAAGAGAGACCCAGGATGATGCTCAAAATACAGAAAG 151  
QY 61 TTACGAAAGGCTTCGATGATTTGCCAAATATCTCTAAGAAAGAGTGGAAAAGATG 120  
Db 152 ATCCAAAAGGCTTCGATGATTTGCCAAATATCTCTAAGAAAGAGTGGAAAAGATG 211  
QY 121 AAATCCTCGAGAAAATCGTCTATGTATATGAAGCTAAATATGAGGTCTAGCTATAA 180  
Db 212 AAAGCCTCGAGAAAATCTCTATGTATATGAAGAAAGATGATGAGGTCTAGCTATAA 271  
QY 181 CTAGGTTTCAAGTCAACCTCCACCTTTCATGCTAGTAAACGGGTGAGACTTCCAC 240  
Db 272 CTAGGTTTCAAGGCAACCTCCACCTTTCATGCTAGTAAACGGGTGAGACTTCCAC 331  
QY 241 GGGAAATGATTTGGTAACGATCGAAACCAAGGATCAGGTTGAACGCTCTCAGATGACT 300  
Db 332 GGGAAATGATTTGGTAACGATCGAAACCAAGGATCAGGTTGAACGCTCTCAGATGACT 391  
QY 301 TTCCGGACGCTCCAGAGAAATCTTCCCGAAGATCATGCCAAGAGCCAGCAGAGGAAGAA 360  
Db 392 TTCCGGACGCTCCAGAGAAATCTTCCCGAAGATCATGCCAAGAGCCAGCAGAGGAAGAA 451  
QY 361 AATGGTTTGAAGGAAGTCCAGAGGCATCTGGCCCAACAATATGGAACACCTGTGC 420  
Db 452 AATGATTCGAGGAAGTGCAGAAAGCATCTGGCCCAACAATATGATGGAAGAGACTGTGC 511  
QY 421 CCCCCGGGAATCCAAGTACCTTGGAGAAGATTAAAGACATCTGGACCCCAAAAGGGG 480  
Db 512 CCCCCGGGAACCAACTCTCTGAGAAGATTACAGAGATCTGGACCCCAAAAGGGG 571

```
481 AAACATGCTCGGACCCACAGACTGCGGTGAGAGAAAGCAGCTGGTGGTTTATGAAGAGATC 540
482 |
483 |
484 |
485 |
486 |
487 |
488 |
489 |
490 |
491 |
492 |
493 |
494 |
495 |
496 |
497 |
498 |
499 |
500 |
501 |
502 |
503 |
504 |
505 |
506 |
507 |
508 |
509 |
510 |
511 |
512 |
513 |
514 |
515 |
516 |
517 |
518 |
519 |
520 |
521 |
522 |
523 |
524 |
525 |
526 |
527 |
528 |
529 |
530 |
531 |
532 |
533 |
534 |
535 |
536 |
537 |
538 |
539 |
540 |
541 |
542 |
543 |
544 |
545 |
546 |
547 |
548 |
549 |
550 |
551 |
552 |
553 |
554 |
555 |
556 |
557 |
558 |
559 |
560 |
561 |
562 |
563 |
564 |
565 |
566 |
567 |
568 |
569 |
570 |
571 |
572 |
573 |
574 |
575 |
576 |
577 |
578 |
579 |
580 |
581 |
582 |
583 |
584 |
585 |
586 |
587 |
588 |
589 |
590 |
591 |
592 |
593 |
594 |
595 |
596 |
597 |
598 |
599 |
600 |
601 |
602 |
603 |
604 |
605 |
606 |
607 |
608 |
609 |
610 |
611 |
612 |
613 |
614 |
615 |
616 |
617 |
618 |
619 |
620 |
621 |
622 |
623 |
624 |
625 |
626 |
627 |
628 |
629 |
630 |
631 |
632 |
633 |
634 |
635 |
636 |
637 |
638 |
639 |
640 |
641 |
642 |
643 |
644 |
645 |
646 |
647 |
648 |
649 |
650 |
651 |
652 |
653 |
654 |
655 |
656 |
657 |
658 |
659 |
660 |
661 |
662 |
663 |
664 |
665 |
666 |
667 |
668 |
669 |
670 |
671 |
672 |
673 |
674 |
675 |
676 |
677 |
678 |
679 |
680 |
681 |
682 |
683 |
684 |
685 |
686 |
687 |
688 |
689 |
690 |
691 |
692 |
693 |
694 |
695 |
696 |
697 |
698 |
699 |
700 |
701 |
702 |
703 |
704 |
705 |
706 |
707 |
708 |
709 |
710 |
711 |
712 |
713 |
714 |
715 |
716 |
717 |
718 |
719 |
720 |
721 |
722 |
723 |
724 |
725 |
726 |
727 |
728 |
729 |
730 |
731 |
732 |
733 |
734 |
735 |
736 |
737 |
738 |
739 |
740 |
741 |
742 |
743 |
744 |
745 |
746 |
747 |
748 |
749 |
750 |
751 |
752 |
753 |
754 |
755 |
756 |
757 |
758 |
759 |
760 |
761 |
762 |
763 |
764 |
765 |
766 |
767 |
768 |
769 |
770 |
771 |
772 |
773 |
774 |
775 |
776 |
777 |
778 |
779 |
780 |
781 |
782 |
783 |
784 |
785 |
786 |
787 |
788 |
789 |
790 |
791 |
792 |
793 |
794 |
795 |
796 |
797 |
798 |
799 |
800 |
801 |
802 |
803 |
804 |
805 |
806 |
807 |
808 |
809 |
810 |
811 |
812 |
813 |
814 |
815 |
816 |
817 |
818 |
819 |
820 |
821 |
822 |
823 |
824 |
825 |
826 |
827 |
828 |
829 |
830 |
831 |
832 |
833 |
834 |
835 |
836 |
837 |
838 |
839 |
840 |
841 |
842 |
843 |
844 |
845 |
846 |
847 |
848 |
849 |
850 |
851 |
852 |
853 |
854 |
855 |
856 |
857 |
858 |
859 |
860 |
861 |
862 |
863 |
864 |
865 |
866 |
867 |
868 |
869 |
870 |
871 |
872 |
873 |
874 |
875 |
876 |
877 |
878 |
879 |
880 |
881 |
882 |
883 |
884 |
885 |
886 |
887 |
888 |
889 |
890 |
891 |
892 |
893 |
894 |
895 |
896 |
897 |
898 |
899 |
900 |
901 |
902 |
903 |
904 |
905 |
906 |
907 |
908 |
909 |
910 |
911 |
912 |
913 |
914 |
915 |
916 |
917 |
918 |
919 |
920 |
921 |
922 |
923 |
924 |
925 |
926 |
927 |
928 |
929 |
930 |
931 |
932 |
933 |
934 |
935 |
936 |
937 |
938 |
939 |
940 |
941 |
942 |
943 |
944 |
945 |
946 |
947 |
948 |
949 |
950 |
951 |
952 |
953 |
954 |
955 |
956 |
957 |
958 |
959 |
960 |
961 |
962 |
963 |
964 |
965 |
966 |
967 |
968 |
969 |
970 |
971 |
972 |
973 |
974 |
975 |
976 |
977 |
978 |
979 |
980 |
981 |
982 |
983 |
984 |
985 |
986 |
987 |
988 |
989 |
990 |
991 |
992 |
993 |
994 |
995 |
996 |
997 |
998 |
999 |
1000 |
```

Db 339 GGGATGATTGGATAATGACCCCTAACCGTGGGAATCAGGTTGAACGCTCCTCAGATGACT 398  
QY 301 TTCGGAGCCTCCAGAGAAATCTTCCGAGATCATGCCAAGAGCCAGCAGAGGAAGAA 360  
Db 399 TTCGGAGGCTCCAGGGAATCTCCCGAAGATCATGCCAAGAGCCAGCAGAGGAAGAA 458  
QY 361 AATGGTTTGAAGGAAGTGCAGAGGATCTGTGGCCCAAAATGATGGGAAACAGCTGTGC 420  
Db 459 AATGATTCGGAGGAAGTGCAGAGGATCTGTGGCCCAAAATGATGGGAAAGAGCTGTGC 518  
QY 421 CCCCGGGAATCCAAGTACCTTGGAGAGATTAACAAGACATCTGGACCCAAAGAGGGG 480  
Db 519 CCCCGGGAATCCAAGTACCTTGGAGAGATTAACAAGACATCTGGACCCAAAGAGGGG 578  
QY 481 AAACATGCTCGGAGCCACACAGATGCTGTGAGAGAAAGCAGCTGTGTATGAAGAGATC 540  
Db 579 GAACATGCTCGGAGCCACACAGATGCTGTGAGAGAAAGCAGCTGTGTATGAAGAGATC 638  
QY 541 AGCAGCCTGAGGAAGATGACGAGTAACCTCCCTC 575  
Db 639 AGCAGCCTGAGGAAGATGACGAGTAACCTCCCTC 673

RESULT 15  
AR060380  
LOCUS AR060380 931 bp DNA linear PAT 29-SEP-1999  
DEFINITION Sequence 2 from patent US 5840568.  
ACCESSION AR060380  
VERSION AR060380.1 GI:5986830  
KEYWORDS  
SOURCE Unknown.  
ORGANISM Unknown.  
REFERENCE 1 (bases 1 to 931)  
AUTHORS Pfreundschuh,M.  
TITLE Hodgkin's disease associated molecules and uses thereof  
JOURNAL Patent: US 5840568-A 2 24-NOV-1998;  
FEATURES  
Location/Qualifiers  
1. .931  
/organism="unknown"  
/mol\_type="unassigned DNA"

ORIGIN  
Query Match 86.2%; Score 496.6; DB 6; Length 931;  
Best Local Similarity 91.5%; Pred. No. 4.7e-140;  
Matches 526; Conservative 0; Mismatches 49; Indels 0; Gaps 0;  
QY 1 ATGAACGGAGACGACGCTTTCGAGGAGACCCAGGATGATGCTCAAAATATCAGAGAAG 60  
Db 99 ATGAACGGAGACGACGCTTTCGAGGAGACCCAGGATGATGCTCAAAATATCAGAGAAG 158  
QY 61 TTACGAAAGCCCTTCGATGATATTGCCAAATACTTCTTAAGAAAGAGTGGGAAAGATG 120  
Db 159 ATCCAAAGCCCTTCGATGATATTGCCAAATACTTCTTAAGAAAGAGTGGGAAAGATG 218  
QY 121 AAATCCTCGGAGAAATCGTCTATGTATATGAAGCTAAACTATGAGGTCATGACTAAA 180  
Db 219 AAAGCCTCGGAGAAATCTTCTATGTATATGAAGAAAGATGAGGCTATGACTAAA 278  
QY 181 CTAGGTTTCAAGGTCACCCCTCCACCTTTCATCGTAGTAAACGGGTCGAGACTTCCAC 240  
Db 279 CTAGGTTTCAAGGTCACCCCTCCACCTTTCATGTGTATTAACCGGCCGAGACTTCCAG 338  
QY 241 GGAATGATTTGGTAAACGATCGAAACACAGAGATCAGGTTGAACGCTCTCAGTACT 300  
Db 339 GGAATGATTTGGTAAACGATCGAAACACAGAGATCAGGTTGAACGCTCTCAGTACT 398  
QY 301 TTCGGAGCCTCCAGAGAAATCTTCCGAGATCATGCCAAGAGCCAGCAGAGGAAGAA 360  
Db 399 TTCGGAGGCTCCAGGGAATCTCCCGAAGATCATGCCAAGAGCCAGCAGAGGAAGAA 458  
QY 361 AATGGTTTGAAGGAAGTGCAGAGGATCTGTGGCCCAAAATGATGGGAAACAGCTGTGC 420  
Db 459 AATGATTCGGAGGAAGTGCAGAGGATCTGTGGCCCAAAATGATGGGAAAGAGCTGTGC 518  
QY 421 CCCCGGGAATCCAAGTACCTTGGAGAGATTAACAAGACATCTGGACCCAAAGAGGGG 480  
Db 519 CCCCGGGAATCCAAGTACCTTGGAGAGATTAACAAGACATCTGGACCCAAAGAGGGG 578  
QY 361 AATGGTTTGAAGGAAGTGCAGAGGATCTGTGGCCCAAAATGATGGGAAACAGCTGTGC 420

Db 459 AATGATTCGGAGGAAGTGCAGAGAATCTGTGGCCCAAAATGATGGGAAAGAGCTGTGC 518  
QY 421 CCCCGGGAATCCAAGTACCTTGGAGAGATTAACAAGACATCTGGACCCAAAGAGGGG 480  
Db 519 CCCCGGGAATCCAAGTACCTTGGAGAGATTAACAAGACATCTGGACCCAAAGAGGGG 578  
QY 481 AAACATGCTCGGAGCCACACAGATGCTGTGAGAGAAAGCAGCTGTGTATGAAGAGATC 540  
Db 579 GAACATGCTCGGAGCCACACAGATGCTGTGAGAGAAAGCAGCTGTGTATGAAGAGATC 638  
QY 541 AGCAGCCTGAGGAAGATGACGAGTAACCTCCCTC 575  
Db 639 AGCAGCCTGAGGAAGATGACGAGTAACCTCCCTC 673

RESULT 15  
AR117874  
LOCUS AR117874 931 bp DNA linear PAT 16-MAY-2001  
DEFINITION Sequence 2 from patent US 6140464.  
ACCESSION AR117874  
VERSION AR117874.1 GI:14098780  
KEYWORDS  
SOURCE Unknown.  
ORGANISM Unknown.  
REFERENCE 1 (bases 1 to 931)  
AUTHORS Pfreundschuh,M. and Rammensee,H.-G.  
TITLE Nonapeptides that bind a HLA-A2.1 molecule  
JOURNAL Patent: US 6140464-A 2 31-OCT-2000;  
FEATURES  
Location/Qualifiers  
1. .931  
/organism="unknown"  
/mol\_type="unassigned DNA"

ORIGIN  
Query Match 86.2%; Score 496.6; DB 6; Length 931;  
Best Local Similarity 91.5%; Pred. No. 4.7e-140;  
Matches 526; Conservative 0; Mismatches 49; Indels 0; Gaps 0;  
QY 1 ATGAACGGAGACGACGCTTTCGAGGAGACCCAGGATGATGCTCAAAATATCAGAGAAG 60  
Db 99 ATGAACGGAGACGACGCTTTCGAGGAGACCCAGGATGATGCTCAAAATATCAGAGAAG 158  
QY 61 TTACGAAAGCCCTTCGATGATATTGCCAAATACTTCTTAAGAAAGAGTGGGAAAGATG 120  
Db 159 ATCCAAAGCCCTTCGATGATATTGCCAAATACTTCTTAAGAAAGAGTGGGAAAGATG 218  
QY 121 AAATCCTCGGAGAAATCGTCTATGTATATGAAGCTAAACTATGAGGTCATGACTAAA 180  
Db 219 AAAGCCTCGGAGAAATCTTCTATGTATATGAAGAAAGATGAGGCTATGACTAAA 278  
QY 181 CTAGGTTTCAAGGTCACCCCTCCACCTTTCATCGTAGTAAACGGGTCGAGACTTCCAC 240  
Db 279 CTAGGTTTCAAGGTCACCCCTCCACCTTTCATGTGTATTAACCGGCCGAGACTTCCAG 338  
QY 241 GGAATGATTTGGTAAACGATCGAAACACAGAGATCAGGTTGAACGCTCTCAGTACT 300  
Db 339 GGAATGATTTGGTAAACGATCGAAACACAGAGATCAGGTTGAACGCTCTCAGTACT 398  
QY 301 TTCGGAGCCTCCAGAGAAATCTTCCGAGATCATGCCAAGAGCCAGCAGAGGAAGAA 360  
Db 399 TTCGGAGGCTCCAGGGAATCTCCCGAAGATCATGCCAAGAGCCAGCAGAGGAAGAA 458  
QY 361 AATGGTTTGAAGGAAGTGCAGAGGATCTGTGGCCCAAAATGATGGGAAACAGCTGTGC 420  
Db 459 AATGATTCGGAGGAAGTGCAGAGGATCTGTGGCCCAAAATGATGGGAAAGAGCTGTGC 518  
QY 421 CCCCGGGAATCCAAGTACCTTGGAGAGATTAACAAGACATCTGGACCCAAAGAGGGG 480  
Db 519 CCCCGGGAATCCAAGTACCTTGGAGAGATTAACAAGACATCTGGACCCAAAGAGGGG 578  
QY 481 AAACATGCTCGGAGCCACACAGATGCTGTGAGAGAAAGCAGCTGTGTATGAAGAGATC 540

Db 579 GAACATGCTGGACCCACAGACTGCGTGAGAGAAAAACAGCTGCTGTAATTATGAAGAGATC 638  
QY 541 AGCGACCCCTGAGGAAGATGACGAGTAACTCCCTC 575  
Db 639 AGCGACCCCTGAGGAAGATGACGAGTAACTCCCTC 673

Search completed: March 31, 2004, 16:39:15  
Job time : 3494 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 31, 2004, 13:52:29 ; Search time 437 Seconds  
(without alignments)  
5599.456 Million cell updates/sec

Title: US-09-975-856-1

Perfect score: 576

Sequence: 1 ATGACGGAGACGACGCGCTT.....ATGACGAGTAACCTCCCTCG 576

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 3373863 seqs, 2124099041 residues

Total number of hits satisfying chosen parameters: 6747726

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

- N\_Geneseq\_29Jan04.\*
- 1: Geneseqn1980s.\*
- 2: Geneseqn1990s.\*
- 3: Geneseqn2000s.\*
- 4: Geneseqn2001as.\*
- 5: Geneseqn2001bs.\*
- 6: Geneseqn2002s.\*
- 7: Geneseqn2003as.\*
- 8: Geneseqn2003bs.\*
- 9: Geneseqn2003cs.\*
- 10: Geneseqn2004s.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	576	100.0	576	6	ABK84472 Human cDN
2	576	100.0	576	7	ABQ83858 Human SSX
3	576	100.0	1250	9	ADC09598 SSX-4 cDN
4	574.4	99.7	576	2	Aav70060 Human SSX
5	496.6	86.2	766	6	Aat11780 Human X-c
6	496.6	86.2	766	6	ABS73282 DNA encod
7	496.6	86.2	766	7	Aad54033 Human col
8	496.6	86.2	766	7	ABQ83844 Human SSX
9	496.6	86.2	766	9	ADC09570 SSX-2 cDN
10	496.6	86.2	766	9	ADD25523 Binding d
11	496.6	86.2	931	2	Aat47748 Melanoma
12	496.6	86.2	931	2	Aav04267 Melanoma
13	496.6	86.2	1309	6	ABL64122 Breast ca
14	496.6	86.2	1309	6	ABL63730 Breast ca
15	487.2	84.6	921	5	AAS80402 DNA encod
16	487.2	84.6	921	5	AAS80399 DNA encod
17	480.6	83.4	576	2	Aav70061 Human SSX
18	480	83.3	766	2	Aat11779 Human X-c
19	221.4	38.4	711	2	Aat11781 Human SYT
20	209.6	36.4	585	2	Aat11782 Human SYT
21	192.6	33.4	830	5	AAS80401 DNA encod
22	191.4	33.2	822	5	AAS80404 DNA encod
23	117.4	20.4	165	6	ABS73281 DNA encod

24	74.2	12.9	1110	5	AAS91978	Aas91978 DNA encod
25	42.2	7.3	2000	7	ADA71938	Ada71938 Rice gene
26	41.4	7.2	1981	9	ADC32171	Adc32171 Human nov
27	40.8	7.1	80	6	ABN33980	Abn33980 Human spl
c 28	40.8	7.1	8059	3	AAA81747	Aaa81747 N. mening
29	40.8	7.1	110000	3	AAA81490_07	Continuation (8 of
30	40.8	7.1	349980	3	AAF21608	Aaf21608 Neisseria
31	39.8	6.9	571	6	ABK71614	Abk71614 Human dit
c 32	39.8	6.9	3591	8	ADA10966	Ada10966 Human cDN
c 33	39.8	6.9	3591	8	ACH04006	Ach04006 Human cDN
34	39.4	6.8	1097	4	AAF27660	Aaf27660 DNA encod
35	39.4	6.8	1097	8	ACH04007	Ach04007 Human cDN
36	37.8	6.6	2736	4	AAI57976	Aai57976 Human pol
37	37.8	6.6	2872	4	AAH98654	Aah98654 Human EST
38	37.8	6.6	2872	4	AAI59762	Aai59762 Human pol
39	37.8	6.6	5532	9	ADB47398	Adb47398 Human cDN
40	37.4	6.5	730	4	AAC91340	Aac91340 Human pol
41	37	6.4	653	6	ABQ99153	Abq99153 Human ORF
42	36	6.2	2369	6	ABA01156	Abao1156 Human zin
43	36	6.2	2798	9	ADC30277	Adc30277 Human nov
c 44	36	6.2	349980	5	AAH41224	Aah41224 Pyrococcu
45	35.8	6.2	765	4	AAK91889	Aak91889 Human cDN

## ALIGNMENTS

RESULT 1  
ABK84472  
ID ABK84472 standard; cDNA; 576 BP.  
XX AC ABK84472;  
DT 14-AUG-2002 (first entry)  
XX DE Human cDNA differentially expressed in granulocytic cells #1043.  
XX KW Human; ss; granulocytic cell; DNA chip; bacterial infection;  
KW viral infection; parasitic infection; protozoal infection;  
KW fungal infection; sterile inflammatory disease; psoriasis;  
KW rheumatoid arthritis; glomerulonephritis; asthma; thrombosis;  
KW cardiac reperfusion injury; renal reperfusion injury; AKDS;  
KW adult respiratory distress syndrome; inflammatory bowel disease;  
KW Crohn's disease; ulcerative colitis; periodontal disease;  
KW granulocyte activation; chronic inflammation; allergy.  
XX OS Homo sapiens.  
XX WO200228999-A2.  
XX 11-APR-2002.  
PD 03-OCT-2001; 2001WO-US030821.  
PF 03-OCT-2000; 2000US-0237189P.  
PR (GENE-) GENE LOGIC INC.  
PA Beazer-Barclay Y, Weissman SM, Yamaga S, Vockley J;  
PI WPI; 2002-435328/46.  
DR Detecting granulocyte activation by detecting differential expression of  
XX genes associated with granulocyte activation, which serves as diagnostic  
XX markers that is useful for monitoring disease states and drug toxicity.  
XX Claim 1; SEQ ID NO 1043; 114pp; English.  
XX The invention relates to detecting (M1) granulocyte (GC) activation  
CC (GCA), by detecting the level of expression of gene(s) (Gs) identified by  
CC DNA chip analysis as given in the specification, and comparing the  
CC expression level to an expression level in an unactivated GC, where  
CC differential expression of Gs is indicative of GCA. Also included are

modulating (M2) GA by contacting GC with an agent that alters the expression of at least one gene in Gs; (2) screening (M3) for an agent capable of modulating GCA or an inflammation (especially chronic) in a tissue, an allergic response in a subject, exposure of a subject to a pathogen or sterile inflammatory disease using the gene expression profile; (3) detecting (M4) an inflammation (especially chronic) in a tissue, an allergic response in a subject, exposure of a subject to a pathogen or sterile inflammatory disease, by detecting the level of expression in a sample of the tissue of gene(s) from Gs, where the level of expression of the gene is indicative of inflammation; (4) treating (M5) an inflammation (especially chronic) or in a tissue, an allergic response in a subject, exposure of a subject to a pathogen or sterile inflammatory disease, by contacting a tissue having inflammation with an agent that modulates the expression of gene(s) from Gs in the tissue. M1 is useful for detecting GCA; M2 is useful for modulating GA; M3 is useful for screening an agent capable of modulating GCA preferably in an inflammation in a tissue; M4 is useful for detecting an inflammation (especially chronic) in a tissue, an allergic response in a subject, exposure of a subject to a pathogen or sterile inflammatory disease (e.g. psoriasis, rheumatoid arthritis, glomerulonephritis, asthma, thrombosis, cardiac reperfusion injury, renal reperfusion injury, ARDS, adult respiratory distress syndrome, inflammatory bowel disease, Crohn's disease, ulcerative colitis, periodontal disease; also bacterial infection, viral infection, parasitic infection, protozoal infection, fungal infection and M5 is useful for treating one of the above conditions. The present sequence represents a gene differentially expressed in granulocytes. Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published\_pat\_sequences

Sequence 576 BP; 187 A; 127 C; 150 G; 112 T; 0 U; 0 Other;

Query Match 100.0%; Score 576; DB 6; Length 576;  
Best Local Similarity 100.0%; Pred. No. 9.3e-172;  
Matches 576; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGAACGGAGAGCGCCCTTTGCAAGGAGAGCCAGGGATGATGCTCAAAATATCAGAGAAG 60  
Db 1 ATGAACGGAGAGCGCCCTTTGCAAGGAGAGCCAGGGATGATGCTCAAAATATCAGAGAAG 60  
QY 61 TTACGAAGGCTTCGATGATATCCCAAAATCTCTCTAGAAAGAGTGGGAAAGATG 120  
Db 61 TTACGAAGGCTTCGATGATATCCCAAAATCTCTCTAGAAAGAGTGGGAAAGATG 120  
QY 121 AAATCCTCGGAGAAATCGTCTATGTGTATATGAAGCTAAACTATGAGTCTATGACTAAA 180  
Db 121 AAATCCTCGGAGAAATCGTCTATGTGTATATGAAGCTAAACTATGAGTCTATGACTAAA 180  
QY 181 CTAGTTTCAAGGTCAACCTCCCACTTTCTATGCTAGTAAACCGGCTGCGAGCTTCCAC 240  
Db 181 CTAGTTTCAAGGTCAACCTCCCACTTTCTATGCTAGTAAACCGGCTGCGAGCTTCCAC 240  
QY 241 GGGATGATTTGGTAAAGCTGAAACACAGGATTCAGTTTGAACGTCCTCAGATGACT 300  
Db 241 GGGATGATTTGGTAAAGCTGAAACACAGGATTCAGTTTGAACGTCCTCAGATGACT 300  
QY 301 TTCGCAGCCTCCAGAGATCTTCCGGAAGATCATGCCCAAGACCCAGCAGAGAGAA 360  
Db 301 TTCGCAGCCTCCAGAGATCTTCCGGAAGATCATGCCCAAGACCCAGCAGAGAGAA 360  
QY 361 AATGTTTGAAGAGAGTCCAGAGGATCTTGCCCAACAAATATGATGGGAAACAGCTGTGC 420  
Db 361 AATGTTTGAAGAGAGTCCAGAGGATCTTGCCCAACAAATATGATGGGAAACAGCTGTGC 420  
QY 421 CCCCGGGGAAATCCAGTACCTTGGAGAGATTAACAGACATCTGGACCCCAAGGGGG 480  
Db 421 CCCCGGGGAAATCCAGTACCTTGGAGAGATTAACAGACATCTGGACCCCAAGGGGG 480  
QY 481 AAACATGCTGACCCCAAGAGCTGCGTGAGAGAAAGAGCTGGTGGTTTATGAAGATC 540  
Db 481 AAACATGCTGACCCCAAGAGCTGCGTGAGAGAAAGAGCTGGTGGTTTATGAAGATC 540

QY 541 AGCGACCTGAGGAGATGACGAGTAACCTCCCTCG 576  
Db 541 AGCGACCTGAGGAGATGACGAGTAACCTCCCTCG 576

## RESULT 2

ABQ83858  
ID ABQ83858 standard; cDNA; 576 BP.  
XX ABQ83858;  
AC ABQ83858;  
XX 03-FEB-2003 (first entry)  
DT Human SSX-4 encoding cDNA SEQ ID NO:599.  
DE Human; epitope; vaccine; immunotherapeutic; cytostatic; immunogenicity;  
XX T cell; gene; ss.  
KW Homo sapiens.  
XX WO200281646-A2.  
PN 17-OCT-2002.  
PD 04-APR-2002; 2002WO-US011101.  
PF 06-APR-2001; 2001US-0282211P.  
XX 07-NOV-2001; 2001US-0337017P.  
PR 07-MAR-2002; 2002US-0363210P.  
XX (CTL1-) CTL IMMUNOTHERAPIES CORP.

XX Simard JUL, Diamond DC, Liu L, Xie Z;  
PI WPI; 2003-067518/06.  
DR P-PSDB; ABP74710.

XX Novel epitopes useful as vaccines, comprises peptides or nucleic acid  
PT encoding the peptides, that are useful epitopes of target-associated  
PT antigens.  
XX Claim 1; Page 185; 352pp; English.

The present invention describes an isolated epitope (I) and an epitope cluster. Also described is a vaccine or immunotherapeutic composition (VC) comprising (I). (I) has cytostatic activity. VC is useful for treating an animal, by administering to an animal the vaccine or immunotherapeutic composition. VC is also useful for evaluating immunogenicity of a vaccine or immunotherapeutic composition, by administering VC to an HLA-transgenic animal and evaluating immunogenicity based on a characteristic of the animal, or by in vitro primary stimulation of a T cell and evaluating immunogenicity. (I) is useful for determining specific T cell frequency, by contacting T cells with a MHC-peptide complex, and further comprises ELISPOT analysis, limiting dilution analysis, flow cytometry, in situ hybridisation and/or polymerase chain reaction (PCR). ABQ83843 to ABQ83858 and ABP74128 to ABP74713 represent sequences used in the exemplification of the present invention

Sequence 576 BP; 187 A; 127 C; 150 G; 112 T; 0 U; 0 Other;

Query Match 100.0%; Score 576; DB 7; Length 576;  
Best Local Similarity 100.0%; Pred. No. 9.3e-172;  
Matches 576; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGAACGGAGAGCGCTTTCGAAGGAGAGCCAGGGATGATGCTCAAAATATCAGAGAAG 60  
Db 1 ATGAACGGAGAGCGCTTTCGAAGGAGAGCCAGGGATGATGCTCAAAATATCAGAGAAG 60  
QY 61 TTACGAAGGCTTCGATGATATTCGCAAAATCTTCTTAAGAAAGAGTGGGAAAGATG 120  
Db 61 TTACGAAGGCTTCGATGATATTCGCAAAATCTTCTTAAGAAAGAGTGGGAAAGATG 120





XX WO9805028-A1.  
 XX 12-NOV-1998.  
 XX 25-FEB-1998; 98WO-US003661.  
 XX 05-MAY-1997; 97US-00851138.  
 XX (LUDW-) LUDWIG INST CANCER RES.  
 XX Gure AO, Tureci O, Sahin U, Tsang S, Scanlan MJ, Knuth A;  
 PI Pfreundschuh M, Old LJ, Chen Y;  
 XX WPI; 1998-610379/51.  
 XX New SSX gene family members - useful for assaying for cancer cells.  
 XX Claim 5; Page 12; 19pp; English.  
 XX This is the nucleotide sequence of a newly isolated human SSX4 cDNA  
 CC clone. The cDNA was isolated from a human testicular cDNA by PCR  
 CC amplification using primers (see AAV70062-63) based on the known SSX2  
 CC sequence. A SSX5 clone (see AAV70061) was also obtained. 2 Forms of SSX4  
 CC were identified. One of these lacked nucleotides 331-466 but was  
 CC otherwise identical to the present SSX4 sequence and is described as an  
 CC alternatively spliced form. SSX4 shares 89.4% homology to SSX1 on the  
 CC nucleotide level and 79.3% at the amino acid level. The inventional  
 CC additionally provides expression vectors, transformed or transfected  
 CC cells that can be used to produce SSX proteins, and primers (see AAV70062  
 CC -73) useful for determining expression of an SSX gene in a sample. The  
 CC new SSX genes can be used to assay for cancers such as melanoma  
 XX  
 SQ Sequence 576 BP; 188 A; 127 C; 149 G; 112 T; 0 U; 0 Other;  
 Query Match 99.7%; Score 574.4; DB 2; Length 576;  
 Best Local Similarity 99.8%; Pred. NO. 3e-171;  
 Matches 575; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
 QY 1 ATGAACGAGACGAGCGCTTTGCAAGAGAGAGCCAGGGATGATGCTCAATATCAGAGAAG 60  
 DB 1 ATGAACGAGAGAGCGCTTTGCAAGAGAGAGCCAGGGATGATGCTCAATATCAGAGAAG 60  
 QY 61 TTACGAAGGCGCTTCGATGATATTGCCAAATACCTTCTAAGAAAGAGTGCGAAAGATG 120  
 DB 61 TTACGAAGGCGCTTCGATGATATTGCCAAATACCTTCTAAGAAAGAGTGCGAAAGATG 120  
 QY 121 AAATCCTCGAGAAATCGTCTATGTATATGAAGCTAACTATGAGGTGATGACTAAA 180  
 DB 121 AAATCCTCGAGAAATCGTCTATGTATATGAAGCTAACTATGAGGTGATGACTAAA 180  
 QY 181 CTAGGTTTCAAGGTCACCGTCCACCTTTTCATGCGTAGTAAACGGGCTGCAGACTTCCAC 240  
 DB 181 CTAGGTTTCAAGGTCACCGTCCACCTTTTCATGCGTAGTAAACGGGCTGCAGACTTCCAC 240  
 QY 241 GGGAAATGATTTGGTAAAGATCGAAGAACACAGGAATCAGGTTGAACGCTCTTCAGATGACT 300  
 DB 241 GGGAAATGATTTGGTAAAGATCGAAGAACACAGGAATCAGGTTGAACGCTCTTCAGATGACT 300  
 QY 301 TTGGGAGCCTCCAGAGAAATCTTCCGGAAGATCATGCCCAAGAGCCGAGAGAGAGAA 360  
 DB 301 TTGGGAGCCTCCAGAGAAATCTTCCGGAAGATCATGCCCAAGAGCCGAGAGAGAGAA 360  
 QY 361 AATGGTTTGAAGGAGTCCAGAGGATCTCGGCCACAAATATGATGGAAACAGCTGTGC 420  
 DB 361 AATGGTTTGAAGGAGTCCAGAGGATCTCGGCCACAAATATGATGGAAACAGCTGTGC 420  
 QY 421 CCCCAGGAAATCCAAAGTACCTTGGAGAGATTAAACAGACATCTGGACCCCAAGGGGG 480  
 DB 421 CCCCAGGAAATCCAAAGTACCTTGGAGAGATTAAACAGACATCTGGACCCCAAGGGGG 480  
 QY 481 AAACATGCTGGACCCACAGACTCGGTGAGAGAAAGAGCTGGTGGTTTATGAAGAGATC 540

DB 481 AAACATGCTGGACCCACAGACTCGGTGAGAGAAAGAGCTGGTGGTTTATGAAGAGATC 540  
 QY 541 AGGACCCCTGAGGAAGATGACGAGTAAGTCCCTCCG 576  
 DB 541 AGGACCCCTGAGGAAGATGACGAGTAAGTCCCTCCG 576  
 RESULT 5  
 AAT11780  
 ID AAT11780 standard; cDNA; 766 BP.  
 XX  
 AC AAT11780;  
 XX  
 DT 16-APR-1996 (first entry)  
 XX  
 DE Human X-chromosome SSX2 cDNA.  
 XX  
 KW Human; SSX2 gene; synovial sarcoma; X-chromosome; breakpoint-2;  
 KW ornithine-delta-aminotransferase; OATL2 gene; Xp11.2; SSX1 gene;  
 KW translocation; chromosome-18; 18q11.2; SYT gene; gene fusion; SYT-SSX2;  
 KW fusion protein; primer; PCR; polymerase chain reaction; Smal; LpsI;  
 KW probe; antibody; monoclonal antibody; humanised antibody; hybridisation;  
 KW antisense; antitumour; recombinant vaccine; vaccinia virus; vector;  
 KW cancer; diagnosis; therapy; ss.  
 XX  
 OS Homo sapiens.  
 XX  
 FH Key Location/Qualifiers  
 FT misc\_difference 230..232  
 FT /\*tag= a  
 FT /codon= seq:TTC, aa:Ser  
 FT misc\_difference 251..253  
 FT /\*tag= b  
 FT /codon= seq:AAG, aa:Asn  
 FT misc\_difference 257..259  
 FT /\*tag= c  
 FT /codon= seq:GAG, aa:Lys  
 FT misc\_difference 377..379  
 FT /\*tag= d  
 FT /codon= seq:CGT, aa:Leu  
 FT misc\_recomb 421..422  
 FT /\*tag= e  
 FT /note= "Breakpoint for SYT translocation"  
 FT misc\_feature 422..766  
 FT /\*tag= f  
 FT /note= "3'-Region present in SYT-SSX2 fusion"  
 FT primer\_binding 747..766  
 FT /\*tag= g  
 FT /note= "Binds primer AAT11784"  
 XX  
 PN WO9602641-A2.  
 XX  
 PD 01-FEB-1996.  
 XX  
 PF 19-JUL-1995; 95WO-GB001704.  
 XX  
 PR 19-JUL-1994; 94GB-00014580.  
 XX  
 PA (CANC-) CANCER RES CAMPAIGN TECHNOLOGY.  
 XX  
 PI Cooper CS, Gusterson BA;  
 XX  
 DR WPI; 1996-105904/11.  
 XX  
 DR P-PSDB; AAR90677.  
 XX  
 XX Diagnosing synovial sarcoma by detection of e.g. SYT and SSX1 sequences,  
 PT based on X18 trans-location - by amplification of trans-location  
 PT sequences or detection of the encoded protein.  
 XX  
 PS Claim 6; Fig 7; 55pp; English.  
 XX  
 CC The sequence represents the SSX2 gene (synovial sarcoma X-chromosome  
 CC breakpoint-2). The gene is at the location of a breakpoint at Xp11.2

CC within an ornithine-delta-aminotransferase OATL2 region, associated with  
 CC the translocation t(X;18)(p11.2;q11.2) found in human synovial sarcomas,  
 CC involving joining of the SVT gene (AAT11778) on chromosome-18 at 18q11.2  
 CC to SSX2. The gene fusion is then transcribed to produce an SVT-SSX2  
 CC (AAT11781) fusion transcript and translated into a fusion protein. SSX2  
 CC may be distinguished from related sequence SSX1 (AAT11779) by digestion  
 CC with SmaI and PstI. Primer AAT11784 has been used in polymerase chain  
 CC reaction amplification of SSX2 sequences. Detection of the gene or its  
 CC product in an abnormal location or as a fusion may be used in diagnosis  
 CC of synovial sarcoma, using primers, probes, humanised antibodies,  
 CC monoclonal antibodies, etc. Antisense oligonucleotides and antibodies may  
 CC also be used therapeutically, and antitumour recombinant vaccines may be  
 CC constructed e.g. in a vaccinia virus vector  
 XX  
 SQ Sequence 766 BP; 229 A; 181 C; 200 G; 156 T; 0 U; 0 Other;

Query Match 86.2%; Score 496.6; DB 2; Length 766;  
 Best Local Similarity 91.5%; Pred. No. 1.5e-146;  
 Matches 526; Conservative 0; Mismatches 49; Indels 0; Gaps 0;  
 QY 1 ATGAACGGAGACGACGCTTTGCAAGGAGACCCAGGATGATGCTCAAAATATCAGAGAAG 60  
 DB 92 ATGACGGAGACGACGCTTTGCAAGGAGACCCAGGATGATGCTCAAAATATCAGAGAAG 151  
 QY 61 TTACGAAAGGCTTCGATGATATTGCCAAATATCTCTTAAGAAAGAGTGGGAAAGATG 120  
 DB 152 ATCAAAAGGCTTCGATGATATTGCCAAATATCTCTTAAGAAAGAGTGGGAAAGATG 211  
 QY 121 AATCTCTCGAGAAATCGTCTATGTATATGAGCTAACTATGAGTCACTGACTTAA 180  
 DB 212 AAGCCCTCGAGAAATCTCTATGTATATGAGGAAAGTATGAGGCTATGACTTAA 271  
 QY 181 CTAGGTTTCAAGGTCACCTCCCAACCTTTCATGCTAGTAAACGGGCTGCAGACTTCCAC 240  
 DB 272 CTAGGTTTCAAGGTCACCTCCCAACCTTTCATGCTAGTAAACGGGCTGCAGACTTCCAG 331  
 QY 241 GGAATGATTTGGTAAAGTTCGAAACACAGGAATCAGGTTGAACCTCTCAGATGACT 300  
 DB 332 GGAATGATTTGGTAAATGACCTTAACCTGCGGGAATCAGGTTGAACCTCTCAGATGACT 391  
 QY 301 TTGGGAGCTTCAGAGAAATCTTCCGAAGATCATGCCAAGAGCCAGCAGAGAGAGAA 360  
 DB 392 TTGGGAGCTTCAGAGAAATCTTCCGAAGATCATGCCAAGAGCCAGCAGAGAGAGAA 451  
 QY 361 AATGTTTGAAGGAAGTGCAGAGGCTCTGGCCCAAAATATGATGGGAAACAGCTCTGC 420  
 DB 452 AATGATTCGAGGAAGTGCAGAGGATCTGGCCCAAAATATGATGGGAAAGAGCTTGC 511  
 QY 421 CCCCCGGGAAATCAAGTACCTTGGAGAAATTAACAGACATCTGGAACCCAAAGGGGG 480  
 DB 512 CCCCCGGGAAATCAAGTACCTTGGAGAAATTAACAGACATCTGGAACCCAAAGGGGG 571  
 QY 481 AAACATCGCTGGACCCACAGACTGCGTGGAGAGAAAGCAGCTGGTGTATGAGAGATC 540  
 DB 572 GAACATCGCTGGACCCACAGACTGCGTGGAGAGAAAGCAGCTGGTGTATGAGAGATC 631  
 QY 541 AGCGACCTGAGGAAGATGACGAGTAACCTCCCTC 575  
 DB 632 AGCGACCTGAGGAAGATGACGAGTAACCTCCCTC 666

RESULT 6  
 ABS73282  
 ID ABS73282 standard; DNA; 766 BP.  
 XX  
 AC ABS73282;  
 XX  
 DT  
 XX  
 XX  
 DE  
 XX  
 KW Chromosome aberration; oncogenic fusion protein; cancer;  
 KW proliferative disease; cellular protein isoform; heat shock protein 90;

KW HSP-90; rheumatoid arthritis; cancer; haematopoietic disorder;  
 KW T cell lymphoma; B cell lymphoma; chronic myeloid leukaemia; CML;  
 KW acute myeloid leukaemia; AML; chronic myelomonocytic leukaemia; CMML;  
 KW acute lymphoblastic leukaemia; ALL; APL; NHL; solid tumour;  
 KW papillary thyroid carcinoma; Ewing's sarcoma; melanoma; liposarcoma;  
 KW rhabdomyosarcoma; synovial sarcoma; viral infection; gene; ds.  
 XX Homo sapiens.  
 XX WO200269900-A2.  
 XX 12-SEP-2002.  
 XX 01-MAR-2002; 2002WO-US006518.  
 XX 01-MAR-2001; 2001US-0272751P.  
 XX (CONF-) CONFORMA THERAPEUTICS CORP.  
 XX Fritz LC, Burrows FJ;  
 XX WPI; 2002-698710/75.  
 XX P-PSDB; ABG95081.  
 XX Treating genetically-defined disease associated with chromosomal  
 PT aberrations yielding oncogenic fusion proteins, e.g. cell proliferative  
 PT diseases, involves administering an inhibitor of heat shock protein 90.  
 XX  
 PS Disclosure; Page 236; 389pp; English.

CC The invention describes a method of treating genetically-defined disease  
 CC associated with chromosomal aberrations yielding oncogenic fusion  
 CC proteins (I), treating cancerous cells containing (I) in a heterogeneous  
 CC cell population, treating proliferative diseases associated with mutant  
 CC protein or cellular protein isoforms (II) dependent on heat shock protein  
 CC (HSP)-90, or selectively treating cells expressing (II) involving  
 CC administering HSP90-inhibitor. The method is useful for treating  
 CC genetically-defined disease with chromosomal aberration yielding  
 CC oncogenic fusion protein, treating cancerous cells containing fusion  
 CC protein in heterogeneous cell population, treating proliferative disease  
 CC (e.g. rheumatoid arthritis or cancer) associated with mutant protein or  
 CC cellular protein isoform dependent on heat shock protein (HSP)-90 (e.g.  
 CC p53), or selectively treating cells expressing mutant protein or cellular  
 CC protein isoform in a patient heterozygous for (II). The method is useful  
 CC for treating a disease e.g. haematopoietic disorder such as T or B cell  
 CC lymphoma, chronic myeloid leukaemia (CML), APL, ALL, NHL and CMML,  
 CC or a disease characterised by a solid tumour such as papillary thyroid  
 CC carcinoma, Ewing's sarcoma, melanoma, liposarcoma, rhabdomyosarcoma and  
 CC synovial sarcoma. The method is also useful for treating viral  
 CC infections. This represents the DNA sequence of a chromosome aberration  
 SQ Sequence 766 BP; 229 A; 181 C; 200 G; 156 T; 0 U; 0 Other;

Query Match 86.2%; Score 496.6; DB 6; Length 766;  
 Best Local Similarity 91.5%; Pred. No. 1.5e-146;  
 Matches 526; Conservative 0; Mismatches 49; Indels 0; Gaps 0;  
 QY 1 ATGAACGGAGACGACGCTTTGCAAGGAGACCCAGGATGATGCTCAAAATATCAGAGAAG 60  
 DB 92 ATGAACGGAGACGACGCTTTGCAAGGAGACCCAGGATGATGCTCAAAATATCAGAGAAG 151  
 QY 61 TTACGAAAGGCTTCGATGATATTGCCAAATATCTCTTAAGAAAGAGTGGGAAAGATG 120  
 DB 152 ATCAAAAGGCTTCGATGATATTGCCAAATATCTCTTAAGAAAGAGTGGGAAAGATG 211  
 QY 121 AATCTCTCGAGAAATCGTCTATGTATATGAGCTAACTATGAGTCACTGACTTAA 180  
 DB 212 AAGCCCTCGAGAAATCTCTATGTATATGAGGAAAGTATGAGGCTATGACTTAA 271  
 QY 181 CTAGGTTTCAAGGTCACCTCCCAACCTTTCATGCTAGTAAACGGGCTGCAGACTTCCAC 240  
 DB 272 CTAGGTTTCAAGGTCACCTCCCAACCTTTCATGCTAGTAAACGGGCTGCAGACTTCCAG 331

QY 241 GGGATGATTTGGTAAAGTCGAAACACAGGAATCAGGTTGAACGTCCTCAGATGACT 300  
DB 332 GGGATGATTTGGATATGACCTAAACCGTGGGAATCAGGTTGAACGTCCTCAGATGACT 391  
QY 301 TTCGGCAGCTCCAGAGAAATCTCCGGAAGATCATGCCAAGAACCCAGCAGAGGAAGAA 360  
DB 392 TTCGGCAGCTCCAGGAATCTCCCGAAGATCATGCCAAGAACCCAGCAGAGGAAGAA 451  
QY 361 AATGTTTGAAGGAAGTCCAGAGGATCTCGGACCAAAATGATGGGAAACAGCTGTGC 420  
DB 452 AATGATTCGGAGGAAGTCCAGAGGATCTCGGACCAAAATGATGGGAAACAGCTGTGC 511  
QY 421 CCCCCGGGAAATCCAAAGTACCTTGAGAGAAATTAACAAGACATCTGGACCCAAAGGGGG 480  
DB 512 CCCCCGGGAAATCCAAAGTACCTTGAGAGAAATTAACAAGACATCTGGACCCAAAGGGGG 571  
QY 481 AAACATGCTCGACCCACAGACTCGTGAGAGAAATTAACAAGACATCTGGAGAGATC 540  
DB 572 GAACATGCTCGACCCACAGACTCGTGAGAGAAATTAACAAGAGATC 631  
QY 541 AGCGACCTCGAGGAAGATGACGAGTAACCTCCCTC 575  
DB 632 AGCGACCTCGAGGAAGATGACGAGTAACCTCCCTC 666

## RESULT 7

AAD54033

ID AAD54033 standard; DNA; 766 BP.

AC AAD54033;

XX AAD54033;

XX AAD54033;

DT 17-JUN-2003 (first entry)

DE Human colon cancer-associated polypeptide gene, SSX-2.

XX Human; colon cancer-associated polypeptide; immune response; therapy;

KW Human; colon cancer-associated polypeptide; immune response; therapy;

KW colon cancer; gene; ds.

XX Homo sapiens.

OS Homo sapiens.

XX WO200290986-A1.

PN 14-NOV-2002.

XX 02-MAY-2002; 2002WO-US013994.

XX 04-MAY-2001; 2001US-00849602.

XX (LUDW-) LUDWIG INST CANCER RES.

PA (CORR ) CORNELL RES FOUND INC.

XX Chen Y, Old LJ, Scanlan MJ, Stockert B;

PI WPI; 2003-112003/10.

XX Diagnosing colon cancer in a subject comprises identifying colon cancer-

PT associated polypeptides as antigens that elicit immune response in colon

PT cancer.

XX Claim 1; Page 90-91; 122pp; English.

PS The invention relates to a method for diagnosing colon cancer in a

XX subject which comprises identifying colon cancer-associated polypeptides

CC as antigens that elicit immune response in colon cancer. The method is

CC useful for diagnosing, determining onset, progression, or regression of

CC colon cancer in a subject, or for selecting a course of treatment of a

CC subject having or suspected of having colon cancer. The colon cancer-

CC associated polypeptides are useful as markers for diagnosing colon

CC cancer, and for following the course of treatment of colon cancer. The

CC present sequence is human colon cancer-associated polypeptide gene

XX Sequence 766 BP; 229 A; 181 C; 200 G; 156 T; 0 U; 0 Other;

SQ

Query Match 86.2%; Score 496.6; DB 7; Length 766;  
Best Local Similarity 91.5%; Pred. No. 1.5e-146;  
Matches 526; Conservative 0; Mismatches 49; Indels 0; Gaps 0;  
QY 1 ATGACCGGACGACGCGCTTTCGAGGAGACCCAGGAGATGCTCAAAATATCAGAGAAG 60  
DB 92 ATGAACCGGAGACGACGCTTTCGAGGAGACCCAGGAGATGCTCAAAATATCAGAGAAG 151  
QY 61 TTACGAAAGGCTTCGAGATGATATGCCAAATACCTTCTTAAGAAAGAGTGGGAAAAGATG 120  
DB 152 ATCCAAAAGGCTTCGAGATGATATGCCAAATACCTTCTTAAGAAAGAGTGGGAAAAGATG 211  
QY 121 AAATCTCGGAGAAAATCGTCTATGTATATGAAGCTAAACTATGAGGTCATGACTAAA 180  
DB 212 AAAGCCTCGGAGAAAATCTTCTATGTATATGAAGAGAAAGTATGAGGCTATGACTAAA 271  
QY 181 CTAGGTTTCAAGGTCACCTCCACCTTTCATGCGTAGTAAACGGGCTGCAGACTTCCAC 240  
DB 272 CTAGGTTTCAAGGTCACCTCCACCTTTCATGCGTAGTAAACGGGCTGCAGACTTCCAC 331  
QY 241 GGGAAATGATTTTGGTAACGATCGAAACACAGGAATCAGGTTGAACGTCCTCAGATGACT 300  
DB 332 GGGAAATGATTTGGTAACGATCGAAACACAGGAATCAGGTTGAACGTCCTCAGATGACT 391  
QY 301 TTCGGCAGCTCCAGAGAAATCTCCCGAAGATCATGCCAAGAACCCAGCAGAGGAAGAA 360  
DB 392 TTCGGCAGCTCCAGGAATCTCCCGAAGATCATGCCAAGAACCCAGCAGAGGAAGAA 451  
QY 361 AATGTTTGAAGGAAGTCCAGAGGATCTCGGACCAAAATGATGGGAAACAGCTGTGC 420  
DB 452 AATGATTCGGAGGAAGTCCAGAGGATCTCGGACCAAAATGATGGGAAACAGCTGTGC 511  
QY 421 CCCCCGGGAAATCCAAAGTACCTTGAGAGAAATTAACAAGACATCTGGACCCAAAGGGGG 480  
DB 512 CCCCCGGGAAATCCAAAGTACCTTGAGAGAAATTAACAAGACATCTGGACCCAAAGGGGG 571  
QY 481 AAACATGCTCGACCCACAGACTCGTGAGAGAAATTAACAAGACATCTGGAGAGATC 540  
DB 572 GAACATGCTCGACCCACAGACTCGTGAGAGAAATTAACAAGAGATC 631  
QY 541 AGCGACCTCGAGGAAGATGACGAGTAACCTCCCTC 575  
DB 632 AGCGACCTCGAGGAAGATGACGAGTAACCTCCCTC 666

## RESULT 8

ABQ83844

ID ABQ83844 standard; cDNA; 766 BP.

XX ABQ83844;

AC ABQ83844;

XX 03-FEB-2003 (first entry)

DT Human SSX-2 encoding cDNA SEQ ID NO:6.

XX Human; epitope; vaccine; immunotherapeutic; cytostatic; immunogenicity;

XX T cell; gene; ss.

XX Homo sapiens.

XX WO200281646-A2.

XX 17-OCT-2002.

XX 04-APR-2002; 2002WO-US011101.

XX 06-APR-2001; 2001US-0282211P.

XX 07-NOV-2001; 2001US-0337017P.

XX 07-MAR-2002; 2002US-0363210P.

XX (CTLI-) CTL IMMUNOTHERAPIES CORP.

XX Simard JLL, Diamond DC, Liu L, Xie Z;

PI



Db 272 CTAGGTTTCAAGGCCACCCCTCCACCTTTCATGTGTAATRAAAGCGGCGAAGACTTCCAG 331  
 Qy 241 GGGAAATGATTTGTAAAGCATCGAAACCAAGGAAATCAGGTTGAACGCTCCTCAGATGACT 300  
 Db 332 GGGAAATGATTTGGGATAATGACCTTAACCGTGGGAATCAGGTTGAACGCTCCTCAGATGACT 391  
 Qy 301 TTCCGGCAGCCTCCAGAGAAATCTCCCGAAGATCATGCCAAGAGCCAGCAGAGGAAGAA 360  
 Db 392 TTCCGGCAGCCTCCAGGAAATCTCCCGAAGATCATGCCAAGAGCCAGCAGAGGAAGAA 451  
 Qy 361 AATGGTTGAAGAGTCCAGAGGCAATCTGGGAGGATTAACAGGATCTGGACCCAAAGGGGG 480  
 Db 452 AATGATTCGGAGGAAGTCCAGAGGCAATCTGGGAGGATTAACAGGATCTGGACCCAAAGGGGG 571  
 Qy 421 CCCCCGGGAAATCCAGTACCTTGGGAGGATTAACAGGATCTGGACCCAAAGGGGG 480  
 Db 512 CCCCCGGGAAATCCAGTACCTTGGGAGGATTAACAGGATCTGGACCCAAAGGGGG 571  
 Qy 481 AAACATGCTCGACCCACAGACTCGTGGAGAGAAAGCAGCTGGTGGTTTATGAAGAGATC 540  
 Db 572 GAACATGCTCGACCCACAGACTCGTGGAGAGAAAGCAGCTGGTGGTTTATGAAGAGATC 631  
 Qy 541 AGCGACCTCGAGAGATGACGAGTAACCTCCCTC 575  
 Db 632 AGCGACCTCGAGAGATGACGAGTAACCTCCCTC 666

RESULT 10  
 ADD25523  
 ID ADD25523 standard; DNA; 766 BP.  
 XX AC ADD25523;  
 XX DT 15-JUN-2004 (first entry)  
 XX DE Binding domain-immunoglobulin fusion protein-associated DNA #46.  
 XX DS ds; Binding domain; immunoglobulin; fusion protein; cytostatic;  
 KW antiarthritic; immunosuppressive; antidiabetic; antithyroid;  
 KW neuroprotective; hinge region; immunoglobulin heavy chain;  
 KW CH2 constant region; CH3 constant region; IgG1;  
 KW antibody dependent cell-mediated cytotoxicity; ADCC; complement fixation;  
 KW malignant condition; B-cell disorder; melanoma; carcinoma; sarcoma;  
 KW rheumatoid arthritis; myasthenia gravis; Grave's disease;  
 KW type I diabetes mellitus; multiple sclerosis; autoimmune disease.  
 XX OS Unidentified.  
 XX XX  
 XX FN US2003118592-A1.  
 XX PD 26-JUN-2003.  
 XX XX  
 XX PF 25-JUL-2002; 2002US-00207655.  
 XX PR 17-JAN-2001; 2001US-0367358P.  
 XX PR 17-JAN-2002; 2002US-00053530.  
 XX PR 03-JUN-2002; 2002US-0385691P.  
 XX XX  
 XX PA (GENE-) GENE-CRAFT INC.  
 XX XX  
 XX PI Ledbetter JA, Hayden-Ledbetter MS, Thompson PA;  
 XX DR WPI; 2003-801317/75.  
 XX XX  
 XX PT New binding domain-immunoglobulin fusion protein, useful for treating a  
 PT subject having or suspected of having a malignant condition or a B-cell  
 XX disorder, e.g. melanoma, Grave's disease or autoimmune disease.  
 XX PS Disclosure; SEQ ID NO 84; 157pp; English.  
 XX XX  
 XX CC Unidentified  
 XX SQ Sequence 766 BP; 229 A; 181 C; 200 G; 156 T; 0 U; 0 Other;

Query Match 86.2%; Score 496.6; DB 9; Length 766;  
 Best Local Similarity 91.5%; Pred. No. 1.5e-146;  
 Matches 526; Conservative 0; Mismatches 49; Indels 0; Gaps 0;

Qy 1 ATGAACGGAGACGACGCTTTTGAAGGAGAGCCAGGAGATGCTCAAAATATCAGAGAAG 60  
 Db 92 ATGAACGGAGACGACGCTTTTGAAGGAGAGCCAGGATGCTCAAAATATCAGAGAAG 151  
 Qy 61 TTACGAAAGCCCTTCGATGATATGCGCAATATCTCTTAAGAAAGATGGGAAAGATG 120  
 Db 152 ATCCAAAGGCTTCGATGATATGCGCAATATCTCTTAAGAAAGATGGGAAAGATG 211  
 Qy 121 AAATCTCCGAGAGAAATCGTCTATGTATATCAAGCTAAATATGAGGTCTGACTTAAA 180  
 Db 212 AAAGCCCTCGAGAAATCTCTATGTATATGAAGAAAGATGAGGTCTGACTTAAA 271  
 Qy 181 CTAGGTTTCAAGGTCACCTCCACCTTTCATCGGTAGTAAACGGGTGACAGCTCCAC 240  
 Db 272 CTAGGTTTCAAGGTCACCTCCACCTTTCATCGGTAGTAAACGGGTGACAGCTCCAC 331  
 Qy 241 GGGAAATGATTTGGTAAAGATCGAAACACAGGAAATCAGGTTGAACGCTCCTCAGATGACT 300  
 Db 332 GGGAAATGATTTGGTAAAGATCGAAACACAGGAAATCAGGTTGAACGCTCCTCAGATGACT 391  
 Qy 301 TTCCGGCAGCCTCCAGAGAAATCTTCCCGAAGATCATGCCAAGAGCCAGCAGAGGAAGAA 360  
 Db 392 TTCCGGCAGCCTCCAGGAAATCTTCCCGAAGATCATGCCAAGAGCCAGCAGAGGAAGAA 451  
 Qy 361 AATGGTTTGAAGAGTCCAGAGGCAATCTGGGAGGATTAACAGGATCTGGACCCAAAGGGGG 480  
 Db 452 AATGATTCGGAGGAAGTCCAGAGGCAATCTGGGAGGATTAACAGGATCTGGACCCAAAGGGGG 571  
 Qy 421 CCCCCGGGAAATCCAGTACCTTGGGAGGATTAACAGGATCTGGACCCAAAGGGGG 480  
 Db 512 CCCCCGGGAAATCCAGTACCTTGGGAGGATTAACAGGATCTGGACCCAAAGGGGG 571  
 Qy 481 AAACATGCTCGACCCACAGACTCGTGGAGAGAAAGCAGCTGGTGGTTTATGAAGAGATC 540  
 Db 572 GAACATGCTCGACCCACAGACTCGTGGAGAGAAAGCAGCTGGTGGTTTATGAAGAGATC 631  
 Qy 541 AGCGACCTCGAGAGATGACGAGTAACCTCCCTC 575  
 Db 632 AGCGACCTCGAGAGATGACGAGTAACCTCCCTC 666

RESULT 11  
 AAT47748  
 ID AAT47748 standard; cDNA; 931 BP.  
 XX AC AAT47748;  
 XX DT 05-JUN-1997 (first entry)  
 XX DE Melanoma cell antigen HOM-MEL 40 cDNA clone.  
 XX KW Melanoma; tumour antigen; HOM-MEL 40; serological fishing; vaccine; ds.  
 XX OS Homo sapiens.  
 XX FN W09640209-A1.  
 XX PD 19-DEC-1996.  
 XX XX  
 XX PF 07-JUN-1996; 96WO-US009726.  
 XX PR 07-JUN-1995; 95US-00479328.  
 XX PR 03-JAN-1996; 96US-00580980.  
 XX PR 10-MAY-1996; 96US-00644116.  
 XX PA (LUDW-) LUDWIG INST CANCER RES.  
 XX XX  
 XX PI Pfreundschuh M, Rammensee H;

```

XX DR 1997-051896/05.
XX DR
XX PT Novel method of serological fishing - allows isolation of molecules (esp.
XX PT antigens) associated with pathological conditions, e.g. Hodgkin's
XX PT disease, cancer or auto-immune disease.
XX PS
XX PS Claim 31; Page 29; 44pp; English.
XX CC
XX CC A cDNA clone (AA747748) codes for a novel melanoma cell antigen
XX CC designated HOM-MEL 40. It was isolated using a novel method in which a
XX CC cDNA library prep'd. from malignant melanoma and expressed in E. coli
XX CC cells. Lysates of the host cells were screened with sera that had been
XX CC treated to remove interfering binding partners. This involved contacting
XX CC the sample with lysates of untransfected host cells and with host cells
XX CC transformed with the same vector (phage lambda) used to make the cDNA
XX CC library. The method, termed serological fishing, can be used to detect
XX CC antigens in human tissues, esp. tumour cells, which are useful in the
XX CC molecular diagnosis of diseases and/or for immunotherapy and gene therapy
XX CC of infectious, autoimmune and malignant diseases (see also AA747747-49).
XX CC HOM-MEL 40 is expressed in 50% of melanomas, 20% of prostate cancers, 20%
XX CC of gastric cancers, 26% of colorectal cancers, 12% of lung cancers and
XX CC 20% of breast cancers, but not in normal tissue. HLA-A2 positive tumour
XX CC cells present a nonamer (see also AA09449- 52) derived from HOM-MEL,
XX CC suggesting that HOM-MEL 40-specific vaccines, useful in inducing
XX CC cytotoxic T lymphocytes, are possible
XX SQ
XX SQ Sequence 931 BP; 274 A; 207 C; 231 G; 219 T; 0 U; 0 Other;

Query Match      86.2%; Score 496.6; DB 2; Length 931;
Best Local Similarity 91.5%; Pred. No. 1.6e-146;
Matches 526; Conservative 0; Mismatches 49; Indels 0; Gaps 0;

Qy 1 ATGACCGGAGCAGCGCTTGCAGGAGAGCCAGGATGATGCTCAAAATATCAGAGAAG 60
Db 99 ATGACCGGAGCAGCGCTTGCAGGAGAGCCAGGATGATGCTCAAAATATCAGAGAAG 158
Qy 61 TTACGAAGGCGCTTCGATGATATTCGCCAAATACATCTCTAAGAAAGAGTGGGAAAGATG 120
Db 159 ATCCAAAGGCGCTTCGATGATATTCGCCAAATACATCTCTAAGAAAGAGTGGGAAAGATG 218
Qy 121 AAATCTCGGAGAAATCGTCTATGTATATGAAGCTAACTATGAGTCTATGACTTAA 180
Db 219 AAAGCTCGGAGAAATCTCTATGTATATGAAGAGAAAGTATGAGGCTATGACTTAA 278
Qy 181 CTAGGTTTCAAGTCAAGTCCAGGATCGAAGGATCGAGTTCAGTTCAGTTCAC 240
Db 279 CTAGGTTTCAAGGACACCTCCACCTTCATGTTATTAAGGAGAAAGTATGAGGCTATGACTTAA 278
Qy 241 GGGATGATTTTGGTAAAGATCGAAGCAGGATCGAGTTCAGTTCAGTTCAGTTCAC 300
Db 339 GGGATGATTTTGGTAAAGATCGAAGCAGGATCGAGTTCAGTTCAGTTCAGTTCAC 398
Qy 301 TTGGGACCTCCAGAGATCTTCCGAGAGATCATGCCCAAGAGCCAGAGAGAGAGAA 360
Db 399 TTGGGACCTCCAGAGATCTTCCGAGAGATCATGCCCAAGAGCCAGAGAGAGAGAA 458
Qy 361 AATGTTTGAAGGAGTCCAGAGGATCTGGCCACAAATATGATGGGAAAGAGCTGTGC 420
Db 459 AATGTTTGAAGGAGTCCAGAGGATCTGGCCACAAATATGATGGGAAAGAGCTGTGC 518
Qy 421 CCCCAGGAAATCCCAAGTACCTTGGAGAGAGATTAACAAGACATCTGGACCCCAAAAGGGGG 480
Db 519 CCCCAGGAAATCCCAAGTACCTTGGAGAGATTAACAAGACATCTGGACCCCAAAAGGGGG 578
Qy 481 AAACATGCTGAGACCAAGATCGTGGTGAAGAGAGAGAGCTGGTGGTTTATGAAGAGATC 540
Db 579 GAACATGCTGAGACCAAGATCGTGGTGAAGAGAGAGAGAGCTGGTGGTTTATGAAGAGATC 638
Qy 541 ACAGACCTTGAGAGAGATGACGAGTAACTCCCTC 575
Db 639 ACAGACCTTGAGAGAGATGACGAGTAACTCCCTC 673

```

```

RESULT 12
AAV04267
ID AAV04267 standard; cDNA; 931 BP.
XX AC AAV04267;
XX DT 22-JUN-1998 (first entry)
XX DE Melanoma antigen clone HOM-MEL-40.
XX KW Melanoma; marker; antigen; diagnosis; serological fishing; human;
XX OS Homo sapiens.
XX PN WO9748721-A1.
XX PD 24-DEC-1997.
XX PF 23-JUN-1997; 97WO-US010926.
XX PR 21-JUN-1996; 96US-00668128.
XX PA (LUDM-) LUDWIG INST CANCER RES.
XX PI Pfreundschuh M;
XX WPI; 1998-063074/06.
XX Nucleic acid and derived protein are markers for Hodgkin's disease - used
XX in identifying immuno-reactive markers of disease.
XX Example 7; Page 30-31; 47pp; English.
XX Clone HOM-MEL-40 was isolated from a human malignant melanoma cDNA
XX library using a method, designated serological fishing, designed to
XX identify immunoreactive markers of disease. In this method, cells
XX characteristic of a disease are used to prepare a cDNA library for
XX transformation of eukaryotic or prokaryotic cells, and the cells grown to
XX express proteins. Patient serum is incubated with the cells used to
XX prepare the library, but not transfected, to remove any components
XX reactive with these cells, then the stripped sample is similarly treated
XX with cells carrying the empty vector. The twice-stripped sample is
XX incubated with lysate of the library cells, so that specific components
XX in the sample may bind to the expression protein. Proteins that form
XX immune complexes are identified as disease markers. In the case of HOM-
XX MEL-40, the new melanoma associated antigen is strongly expressed in
XX melanoma, but not healthy tissues. The deduced amino acid sequence
XX includes 3 tumour-associated peptides (see AA41587-89) that bind to HLA-
XX A2.1. The serological fishing method was also used to identify a claimed
XX marker (see AAV04262) of Hodgkin's disease
XX SQ
XX SQ Sequence 931 BP; 274 A; 207 C; 231 G; 219 T; 0 U; 0 Other;

Query Match      86.2%; Score 496.6; DB 2; Length 931;
Best Local Similarity 91.5%; Pred. No. 1.6e-146;
Matches 526; Conservative 0; Mismatches 49; Indels 0; Gaps 0;

Qy 1 ATGACCGGAGCAGCGCTTGCAGGAGAGCCAGGATGATGCTCAAAATATCAGAGAAG 60
Db 99 ATGACCGGAGCAGCGCTTGCAGGAGAGCCAGGATGATGCTCAAAATATCAGAGAAG 158
Qy 61 TTACGAAGGCGCTTCGATGATATTCGCCAAATACATCTCTAAGAAAGAGTGGGAAAGATG 120
Db 159 ATCCAAAGGCGCTTCGATGATATTCGCCAAATACATCTCTAAGAAAGAGTGGGAAAGATG 218
Qy 121 AAATCTCGGAGAAATCGTCTATGTATATGAAGCTAACTATGAGTCTATGACTTAA 180
Db 219 AAAGCTCGGAGAAATCTCTATGTATATGAAGAGAAAGTATGAGGCTATGACTTAA 278
Qy 181 CTAGGTTTCAAGTCAAGTCCAGGATCGAAGGATCGAGTTCAGTTCAGTTCAC 240

```



Db 279 CTAGTTTCAAGGCCACCCCTCCACCTTTTCATGTGTATTAACACGGGCGGAGACTTCCAG 338  
 Qy 241 GGGATGATTGTTGTAACGATCGAAACACACAGGATCAGGTGAACTCTTCAAGTACT 300  
 Db 339 GGGATGATTGTTGTAATGATACCTTAAACCGTGGGATCAGGTGAACTCTTCAAGTACT 398  
 Qy 301 TTGGGACGCTCCAGAGATCTTCCCGAAGATCATGCCCAAGAGCCAGCAGAGGAAGAA 360  
 Db 399 TTGGGACGCTCCAGGGAATCTCCCGAAGATCATGCCCAAGAGCCAGCAGAGGAAGAA 458  
 Qy 361 AATGTTTGAAGAGTCCAGAGGATCTGGCCCAACAAATGATGGAAACAGCTTGC 420  
 Db 459 AATGTTTGAAGAGTCCAGAGGATCTGGCCCAACAAATGATGGAAACAGCTTGC 518  
 Qy 421 CCCCCGGGAATCAAGTACCTTCGAGAGATTAACAAAGATCTGGACCCAAAGGGG 480  
 Db 519 CCCCCGGGAACCAATCTCTGAGAGATTAACAAAGATCTGGACCCAAAGGGG 578  
 Qy 481 AAACATGCTGGACCCACAGACTCGGTGAGAGAAAGAGCTGGTGTATGAGAGATC 540  
 Db 579 GAACATGCTGGACCCACAGACTCGGTGAGAGAAACAGCTGGTGTATGAGAGATC 638  
 Qy 541 AGGACCTGAGGAGATGAGGATTAACCTCCCTC 575  
 Db 639 AGGACCTGAGGAGATGAGGATTAACCTCCCTC 673  
 RESULT 13  
 ABL64122  
 ID ABL64122 standard; DNA; 1309 BP.  
 XX ABL64122;  
 XX  
 XX 15-MAY-2002 (first entry)  
 XX  
 XX Breast cancer related gene sequence SEQ ID NO:2459.  
 XX Human; cancer; colon; breast; ovary; oesophagus; kidney; thyroid;  
 KW stomach; lung; prostate; pancreas; carcinoma; antitumour; cancerous;  
 KW cytosolic; gene therapy; antineoplastic; Wilm's tumour; adenocarcinoma;  
 KW gene; ds.  
 XX  
 XX Homo sapiens.  
 XX  
 XX WO200194629-A2.  
 XX  
 XX 13-DEC-2001.  
 XX  
 XX 30-MAY-2001; 2001WO-US010838.  
 XX  
 XX 05-JUN-2000; 2000US-0209473P.  
 XX 05-JUN-2000; 2000US-0209531P.  
 XX 18-SEP-2000; 2000US-0233133P.  
 XX 18-SEP-2000; 2000US-0233617P.  
 XX 20-SEP-2000; 2000US-0234009P.  
 XX 20-SEP-2000; 2000US-0234034P.  
 XX 20-SEP-2000; 2000US-0234052P.  
 XX 22-SEP-2000; 2000US-0234509P.  
 XX 22-SEP-2000; 2000US-0234567P.  
 XX 25-SEP-2000; 2000US-0234923P.  
 XX 25-SEP-2000; 2000US-0234924P.  
 XX 25-SEP-2000; 2000US-0235077P.  
 XX 25-SEP-2000; 2000US-0235082P.  
 XX 25-SEP-2000; 2000US-0235134P.  
 XX 25-SEP-2000; 2000US-0235280P.  
 XX 26-SEP-2000; 2000US-0235637P.  
 XX 26-SEP-2000; 2000US-0235638P.  
 XX 26-SEP-2000; 2000US-0235711P.  
 XX 27-SEP-2000; 2000US-0235720P.  
 XX 27-SEP-2000; 2000US-0235840P.  
 XX 27-SEP-2000; 2000US-0235863P.  
 XX 28-SEP-2000; 2000US-0236028P.  
 XX 28-SEP-2000; 2000US-0236032P.

PR 28-SEP-2000; 2000US-0236033P.  
 PR 28-SEP-2000; 2000US-0236034P.  
 PR 28-SEP-2000; 2000US-0236109P.  
 PR 28-SEP-2000; 2000US-0236111P.  
 PR 29-SEP-2000; 2000US-0236842P.  
 PR 29-SEP-2000; 2000US-0236891P.  
 PR 02-OCT-2000; 2000US-0237172P.  
 PR 02-OCT-2000; 2000US-0237173P.  
 PR 02-OCT-2000; 2000US-0237278P.  
 PR 02-OCT-2000; 2000US-0237294P.  
 PR 02-OCT-2000; 2000US-0237295P.  
 PR 02-OCT-2000; 2000US-0237316P.  
 PR 03-OCT-2000; 2000US-0237425P.  
 PR 03-OCT-2000; 2000US-0237598P.  
 PR 03-OCT-2000; 2000US-0237604P.  
 PR 03-OCT-2000; 2000US-0237606P.  
 PR 03-OCT-2000; 2000US-0237608P.  
 PR 01-NOV-2000; 2000US-0244867P.  
 PR 01-NOV-2000; 2000US-0245084P.  
 PA (AVAL-) AVALON PHARM.  
 XX  
 XX Young PE, Augustus M, Carter KC, Ebner R, Endress G, Horrigan S;  
 PI Soppet DR, Weaver Z;  
 XX  
 XX WPI; 2002-188264/24.  
 XX  
 XX Screening for anti-neoplastic agent involves exposing cells to a chemical  
 PT agent to be tested for anti-neoplastic activity, and determining a change  
 PT in expression of a gene of a signature gene set.  
 XX  
 XX Claim 1; SEQ ID NO 2459; 44pp; English.  
 XX  
 XX The present invention describes a method (M1) for screening for an anti-  
 CC neoplastic agent. The method involves exposing cells to a chemical agent  
 CC to be tested for anti-neoplastic activity, determining a change in  
 CC expression of at least one gene (I) of a signature gene set, where (I)  
 CC comprises a sequence (S) selected from 8447 sequences (given in ABL61664  
 CC to ABL70110), or is at least 95% identical to (S), where a change in  
 CC expression is indicative of anti-neoplastic activity. (I) has cytostatic  
 CC activity and can be used in gene therapy. M1 can be used for screening an  
 CC anti-neoplastic agent, and can be used for producing a product which is  
 CC the data collected with respect to the anti-neoplastic agent as a result  
 CC of M1, and the data is sufficient to convey the chemical structure and/or  
 CC properties of the agent. M1 can be used in the treatment of cancer such  
 CC as colon, breast, stomach, lung, thyroid, oesophageal, ovarian, kidney,  
 CC prostate or pancreatic cancer, adenocarcinoma, carcinoma, clear cell  
 CC cancer, infiltrating ductal cancer, infiltrating lobular cancer, squamous  
 CC cell carcinoma, neuroendocrine carcinoma, papillary carcinoma and Wilm's  
 CC tumour  
 XX  
 XX Sequence 1309 BP; 385 A; 329 C; 281 G; 309 T; 0 U; 5 Other;  
 Qy Query Match 86.2%; Score 496.6; DB 6; Length 1309;  
 Best Local Similarity 91.5%; Pred. No. 1.9e-146;  
 Matches 526; Conservative 0; Mismatches 49; Indels 0; Gaps 0;  
 Qy 1 ATGAACGGAGACGACGCGCTTTGCAAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 60  
 Db 99 ATGAACGGAGACGACGCGCTTTGCAAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 158  
 Qy 61 TTACGAAGGCGCTTCGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 120  
 Db 159 ATCCAAAAGGCGCTTCGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 218  
 Qy 121 AAATCTCTCGGAGAAAATCGTCTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 180  
 Db 219 AAGCCCTCGGAGAAAATCTTCTAIGTATGATGATGATGATGATGATGATGATGATGATGATGATGATG 278  
 Qy 181 CTAGTTTCAAGGTCACCCCTCCACCTTTCATCGTAGTAGTAAACGGCTGAGACTTCCAC 240  
 Db 279 CTAGTTTCAAGGTCACCCCTCCACCTTTCATCGTAGTAGTAAACGGCTGAGACTTCCAC 338



QY 241 GGAATGATTTTGGTAACGATGAAACACAGGAATCAGGTTGAAAGTCTCTCAGATGACT 300  
 Db 339 GGAATGATTTGATATGACCTTAACCGTGGAAATCAGGTTGAAAGTCTCTCAGATGACT 398  
 QY 301 TTGGGAGGCTCCAGAGAATCTTCCCGAAGATCATGCCAAGAAGCCAGAGGAAGAA 360  
 Db 399 TTGGGAGGCTCCAGGAATCTCCCGAAGATCATGCCAAGAAGCCAGAGGAAGGA 458  
 QY 361 AATGTTTGAAGAAGTCCAGAGGATCTGGCCCAAGATGATGGGAACAGCTGTGC 420  
 Db 459 AATGATTCGGAGAAGTCCAGAGGATCTGGCCCAAGATGATGGGAACAGCTGTGC 518  
 QY 421 CCCCCGGGAATCCAAGTACCTTTGGAGAAGATTAAAGACATCTGGACCCAAAGGGG 480  
 Db 519 CCCCCGGGAATCCAAGTACCTTTGGAGAAGATTAAAGACATCTGGACCCAAAGGGG 578  
 QY 481 AATGATGCTGGACCCAGAGTCTGGCCCAAGATGATGGGAACAGCTGTGC 540  
 Db 579 GAACATGCTGGACCCAGAGTCTGGCCCAAGATGATGGGAACAGCTGTGC 638  
 QY 541 AGCGACCTCAGGAAGTACGAGTAACTCCCTC 575  
 Db 639 AGCGACCTCAGGAAGTACGAGTAACTCCCTC 673  
 RESULT 14  
 ABL63730  
 ID ABL63730 standard; DNA; 1309 BP.  
 XX AC ABL63730;  
 XX XX  
 DT 15-MAY-2002 (first entry)  
 DE Breast cancer related gene sequence SEQ ID NO:2067.  
 XX  
 KW Human; cancer; colon; breast; ovary; oesophagus; kidney; thyroid;  
 KW stomach; lung; prostate; pancreas; carcinoma; antitumour; cancerous;  
 KW cytostatic; gene therapy; antineoplastic; Wilm's tumour; adenocarcinoma;  
 KW gene; ds.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WC200194629-A2.  
 XX  
 PD 13-DEC-2001.  
 XX  
 PF 30-MAY-2001; 2001WO-US010838.  
 XX  
 PR 05-JUN-2000; 2000US-0209473P.  
 PR 05-JUN-2000; 2000US-0209531P.  
 PR 18-SEP-2000; 2000US-0233133P.  
 PR 18-SEP-2000; 2000US-0233617P.  
 PR 20-SEP-2000; 2000US-0234009P.  
 PR 20-SEP-2000; 2000US-0234034P.  
 PR 20-SEP-2000; 2000US-0234052P.  
 PR 22-SEP-2000; 2000US-0234509P.  
 PR 22-SEP-2000; 2000US-0234567P.  
 PR 25-SEP-2000; 2000US-0234923P.  
 PR 25-SEP-2000; 2000US-0234924P.  
 PR 25-SEP-2000; 2000US-0235077P.  
 PR 25-SEP-2000; 2000US-0235082P.  
 PR 25-SEP-2000; 2000US-0235134P.  
 PR 25-SEP-2000; 2000US-0235280P.  
 PR 26-SEP-2000; 2000US-0235637P.  
 PR 26-SEP-2000; 2000US-0235638P.  
 PR 27-SEP-2000; 2000US-0235711P.  
 PR 27-SEP-2000; 2000US-0235720P.  
 PR 27-SEP-2000; 2000US-0235840P.  
 PR 27-SEP-2000; 2000US-0235863P.  
 PR 28-SEP-2000; 2000US-0236028P.  
 PR 28-SEP-2000; 2000US-0236032P.  
 PR 28-SEP-2000; 2000US-0236033P.  
 PR 28-SEP-2000; 2000US-0236034P.

PR 28-SEP-2000; 2000US-0236109P.  
 PR 28-SEP-2000; 2000US-0236111P.  
 PR 29-SEP-2000; 2000US-0236842P.  
 PR 29-SEP-2000; 2000US-0236891P.  
 PR 02-OCT-2000; 2000US-0237172P.  
 PR 02-OCT-2000; 2000US-0237173P.  
 PR 02-OCT-2000; 2000US-0237278P.  
 PR 02-OCT-2000; 2000US-0237294P.  
 PR 02-OCT-2000; 2000US-0237295P.  
 PR 02-OCT-2000; 2000US-0237316P.  
 PR 03-OCT-2000; 2000US-0237425P.  
 PR 03-OCT-2000; 2000US-0237598P.  
 PR 03-OCT-2000; 2000US-0237604P.  
 PR 03-OCT-2000; 2000US-0237606P.  
 PR 01-NOV-2000; 2000US-0244867P.  
 PR 01-NOV-2000; 2000US-0245084P.  
 XX  
 PA (AVAL-) AVALON PHARM.  
 XX  
 PI Young PE, Augustus M, Carter KC, Ebner R, Endress G, Horrigan S;  
 PI Soppet DR, Weaver Z;  
 XX  
 XX WPI; 2002-188264/24.  
 XX  
 XX  
 PT Screening for anti-neoplastic agent involves exposing cells to a chemical  
 PT agent to be tested for anti-neoplastic activity, and determining a change  
 PT in expression of a gene of a signature gene set.  
 XX  
 PS Claim 1; SEQ ID NO 2067; 44pp; English.  
 XX  
 CC The present invention describes a method (M1) for screening for an anti-  
 CC neoplastic agent. The method involves exposing cells to a chemical agent  
 CC to be tested for anti-neoplastic activity, determining a change in  
 CC expression of at least one gene (I) of a signature gene set, where (I)  
 CC comprises a sequence (S) selected from 8447 sequences (given in ABL61664  
 CC to ABL70110), or is at least 95% identical to (S), where a change in  
 CC expression is indicative of anti-neoplastic activity. (I) has cytostatic  
 CC activity and can be used in gene therapy. M1 can be used for screening an  
 CC anti-neoplastic agent, and can be used for producing a product which is  
 CC the data collected with respect to the anti-neoplastic agent as a result  
 CC of M1, and the data is sufficient to convey the chemical structure and/or  
 CC properties of the agent. M1 can be used in the treatment of cancer such  
 CC as colon, breast, stomach, lung, thyroid, oesophageal, ovarian, kidney,  
 CC prostate or pancreatic cancer, adenocarcinoma, carcinoma, clear cell  
 CC cancer, infiltrating ductal cancer, infiltrating lobular cancer, squamous  
 CC cell carcinoma, neuroendocrine carcinoma, papillary carcinoma and Wilm's  
 CC tumour  
 XX  
 SQ Sequence 1309 BP; 385 A; 329 C; 281 G; 309 T; 0 U; 5 Other;  
 XX  
 Query Match 86.2%; Score 496.6; DB 6; Length 1309;  
 Best Local Similarity 91.5%; Pred. No. 1.9e-146;  
 Matches 526; Conservative 0; Mismatches 49; Indels 0; Gaps 0;  
 QY 1 ATGACGGAGAGACGACGCTTTGCAAGGAGAGCCAGGATGATGCTCAATATCAGAGAAG 60  
 Db 99 ATGACGGAGAGACGACGCTTTGCAAGGAGAGCCAGGATGATGCTCAATATCAGAGAAG 158  
 QY 61 TTACGAAAGGCTTCGATGATATGTCCTCAATATGTCCTTAAGAAAGAGTGGGAAAGATG 120  
 Db 159 ATCCAAAGGCTTCGATGATATGTCCTCAATATGTCCTTAAGAAAGAGTGGGAAAGATG 218  
 QY 121 AAATCTCTCGAGAAATCGTCTATGTTATATGAAGCTAACTATGAGGTCTGACTTAA 180  
 Db 219 AAAGCTCGAGAAATCTTCTATGTTATGAAGAAAGATGATGAGGTCTGACTTAA 278  
 QY 181 CTAGGTTTCAAGTACCCCTCCACCTTTCATCGGTAGTAAACGGGTGCGAGCTTCCAC 240  
 Db 279 CTAGGTTTCAAGGCCACCCCTCCACCTTTCATGTTAATAAACGGGCCGAGACTTCCAG 338  
 QY 241 GGAATGATTTTGGTAAACGATCGAAACACAGGAATCAGGTTGAAAGTCTCTCAGATGACT 300







```

; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-09-796-780-1

Query Match      100.0%; Score 576; DB 4; Length 576;
Best Local Similarity 100.0%; Pred. No. 1.7e-184;
Matches 576; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGAACGAGACGAGCGCTTTGCAAGGAGAGCCAGGGATGATGCTCAAAATATCAGAGAG 60
Db 1 ATGAACGAGACGAGCGCTTTGCAAGGAGAGCCAGGGATGATGCTCAAAATATCAGAGAG 60
QY 61 TTACGAAAGGCGCTTCGATGATATTCGCAAAATATCTCTAAGAAAGAGTGGGAAAGATG 120
Db 61 TTACGAAAGGCGCTTCGATGATATTCGCAAAATATCTCTAAGAAAGAGTGGGAAAGATG 120
QY 121 AAATCCTCGGAGAAATCGTCTATGTGATATGAGCTAAACTATGAGGTGATGACTAAA 180
Db 121 AAATCCTCGGAGAAATCGTCTATGTGATATGAGCTAAACTATGAGGTGATGACTAAA 180
QY 181 CTAGGTTCAAGGTCACCTCCACCTTTTCATGGTAGTAAACGGGCTGCAGACTTCCAC 240
Db 181 CTAGGTTCAAGGTCACCTCCACCTTTTCATGGTAGTAAACGGGCTGCAGACTTCCAC 240
QY 241 GGGAAATGATTTTGGTAACGATCGAATCAGAGGATCAGGTTGAACGCTCCTCAGATGACT 300
Db 241 GGGAAATGATTTTGGTAACGATCGAATCAGAGGATCAGGTTGAACGCTCCTCAGATGACT 300
QY 301 TTCCGACGCTCCAGAGAAATCTCCGAAAGATCATGCCAAGAGCCAGAGAGAGAA 360
Db 301 TTCCGACGCTCCAGAGAAATCTCCGAAAGATCATGCCAAGAGCCAGAGAGAGAA 360
QY 361 AATGTTTGAAGGAGTCCAGAGGATCTGGAGAGATTAACAGACATCTGGACCCAAAGGGGG 420
Db 361 AATGTTTGAAGGAGTCCAGAGGATCTGGAGAGATTAACAGACATCTGGACCCAAAGGGGG 420
QY 421 CCCCCGGGAAATCCAAGTACCTTGGAGAGATTAACAGACATCTGGACCCAAAGGGGG 480
Db 421 CCCCCGGGAAATCCAAGTACCTTGGAGAGATTAACAGACATCTGGACCCAAAGGGGG 480
QY 481 AAACATGCTGGACCCAGAGTCCGAGAGAGTCCGAGAGAGAGAGAGATC 540
Db 481 AAACATGCTGGACCCAGAGTCCGAGAGAGTCCGAGAGAGAGAGATC 540
QY 541 AGCGACCTGAGGAGAGATGACGAGTAACTCCCTCTCG 576
Db 541 AGCGACCTGAGGAGAGATGACGAGTAACTCCCTCTCG 576

RESULT 4
US-09-344-040C-5
; Sequence 5, Application US/09344040C
; Patent No. 6548064
; GENERAL INFORMATION:
; APPLICANT: Tureci, Ozlem
; APPLICANT: Sahin, Ugur
; APPLICANT: Pfreundschuh, Michael
; APPLICANT: Ramensee, Hans Georg
; APPLICANT: Stevanovic, Stefan
; TITLE OF INVENTION: Method for Determining Presence of Cancer in a Sample By Determin
; TITLE OF INVENTION: Expression of an SSX Gene, Peptides Derived From Said SSX Gene a
; FILE REFERENCE: LUD 5556.1
; CURRENT APPLICATION NUMBER: US/09/344,040C
; PRIOR FILING DATE: 1999-06-25
; PRIOR APPLICATION NUMBER: US 09/105,839
; PRIOR FILING DATE: 1998-06-26
; PRIOR APPLICATION NUMBER: US 08/851,130
; PRIOR FILING DATE: 1997-05-05
; NUMBER OF SEQ ID NOS: 132
; SEQ ID NO 5
; LENGTH: 576
; TYPE: DNA
; ORGANISM: Homo sapiens

```

```

; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-344-040C-5

Query Match      100.0%; Score 576; DB 4; Length 576;
Best Local Similarity 100.0%; Pred. No. 1.7e-184;
Matches 576; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGAACGAGACGAGCGCTTTGCAAGGAGAGCCAGGGATGATGCTCAAAATATCAGAGAG 60
Db 1 ATGAACGAGACGAGCGCTTTGCAAGGAGAGCCAGGGATGATGCTCAAAATATCAGAGAG 60
QY 61 TTACGAAAGGCGCTTCGATGATATTCGCAAAATATCTCTAAGAAAGAGTGGGAAAGATG 120
Db 61 TTACGAAAGGCGCTTCGATGATATTCGCAAAATATCTCTAAGAAAGAGTGGGAAAGATG 120
QY 121 AAATCCTCGGAGAAATCGTCTATGTGATATGAGCTAAACTATGAGGTGATGACTAAA 180
Db 121 AAATCCTCGGAGAAATCGTCTATGTGATATGAGCTAAACTATGAGGTGATGACTAAA 180
QY 181 CTAGGTTCAAGGTCACCTCCACCTTTTCATGGTAGTAAACGGGCTGCAGACTTCCAC 240
Db 181 CTAGGTTCAAGGTCACCTCCACCTTTTCATGGTAGTAAACGGGCTGCAGACTTCCAC 240
QY 241 GGGAAATGATTTTGGTAACGATCGAATCAGAGGATCAGGTTGAACGCTCCTCAGATGACT 300
Db 241 GGGAAATGATTTTGGTAACGATCGAATCAGAGGATCAGGTTGAACGCTCCTCAGATGACT 300
QY 301 TTCCGACGCTCCAGAGAAATCTCCGAAAGATCATGCCAAGAGCCAGAGAGAGAA 360
Db 301 TTCCGACGCTCCAGAGAAATCTCCGAAAGATCATGCCAAGAGCCAGAGAGAGAA 360
QY 361 AATGTTTGAAGGAGTCCAGAGGATCTGGAGAGATTAACAGACATCTGGACCCAAAGGGGG 420
Db 361 AATGTTTGAAGGAGTCCAGAGGATCTGGAGAGATTAACAGACATCTGGACCCAAAGGGGG 420
QY 421 CCCCCGGGAAATCCAAGTACCTTGGAGAGATTAACAGACATCTGGACCCAAAGGGGG 480
Db 421 CCCCCGGGAAATCCAAGTACCTTGGAGAGATTAACAGACATCTGGACCCAAAGGGGG 480
QY 481 AAACATGCTGGACCCAGAGTCCGAGAGAGTCCGAGAGAGAGAGATC 540
Db 481 AAACATGCTGGACCCAGAGTCCGAGAGAGTCCGAGAGAGAGAGATC 540
QY 541 AGCGACCTGAGGAGAGATGACGAGTAACTCCCTCTCG 576
Db 541 AGCGACCTGAGGAGAGATGACGAGTAACTCCCTCTCG 576

RESULT 5
US-09-833-039A-5
; Sequence 5, Application US/09833039A
; Patent No. 6673350
; GENERAL INFORMATION:
; APPLICANT: Tureci, Ozlem
; APPLICANT: Sahin, Ugur
; APPLICANT: Pfreundschuh, Michael
; TITLE OF INVENTION: Tumor Associated Peptide and Uses Thereof
; FILE REFERENCE: LUD 5622.1
; CURRENT APPLICATION NUMBER: US/09/833,039A
; PRIOR FILING DATE: 2001-04-12
; PRIOR APPLICATION NUMBER: US 09/409,455
; PRIOR FILING DATE: 1999-06-25
; PRIOR APPLICATION NUMBER: US 09/344,040
; PRIOR FILING DATE: 1999-06-26
; PRIOR APPLICATION NUMBER: US 09/105,839
; PRIOR FILING DATE: 1998-06-26
; NUMBER OF SEQ ID NOS: 131
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 5
; LENGTH: 576
; TYPE: DNA
; ORGANISM: Homo sapiens

```

US-09-833-039A-5

Query Match 100.0%; Score 576; DB 4; Length 576;  
 Best Local Similarity 100.0%; Pred. No. 1.7e-184;  
 Matches 576; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ATGAACGGAGACGACGCTTTGCAAGGAGAGACCCAGGGATGATGCTCAAAATATCAGAGAAG 60  
 Db 1 ATGAACGGAGACGACGCTTTGCAAGGAGAGACCCAGGGATGATGCTCAAAATATCAGAGAAG 60

Qy 61 TTACGAAGGCTTCGATGATATTGCCAAATACCTTCTTAAGAAAGAGTGGAAAGATG 120  
 Db 61 TTACGAAGGCTTCGATGATATTGCCAAATACCTTCTTAAGAAAGAGTGGAAAGATG 120

Qy 121 AAATCCTCGGAGAAATCGTCTATGTATATGAAGCTAAACTATGAGGTCACTATAA 180  
 Db 121 AAATCCTCGGAGAAATCGTCTATGTATATGAAGCTAAACTATGAGGTCACTATAA 180

Qy 181 CTAGGTTTCAAGGTCACTCCACCTTTCATGCGTAGTAAACCGGCTGCAGACTTCCAC 240  
 Db 181 CTAGGTTTCAAGGTCACTCCACCTTTCATGCGTAGTAAACCGGCTGCAGACTTCCAC 240

Qy 241 GGGATGATTTTGGTAACGATCGAAACCAACAGGAATCAGGTTGAACGTCCTCAGTACT 300  
 Db 241 GGGATGATTTTGGTAACGATCGAAACCAACAGGAATCAGGTTGAACGTCCTCAGTACT 300

Qy 301 TTCCGACGCTCCAGAGAAATCTTCCGAAAGATCATGCCAAGAGCCAGCAGAGAGAGAA 360  
 Db 301 TTCCGACGCTCCAGAGAAATCTTCCGAAAGATCATGCCAAGAGCCAGCAGAGAGAGAA 360

Qy 361 AATGGTTGAAGGAGTCCAGAGCATCTCGGCCCAAAATGATGGAAACAGCTGTGC 420  
 Db 361 AATGGTTGAAGGAGTCCAGAGCATCTCGGCCCAAAATGATGGAAACAGCTGTGC 420

Qy 421 CCCCCGGGAAATCAAAGTACCTTGGAGAGATTAACAGAGATCTGGACCCCAAGAGGGG 480  
 Db 421 CCCCCGGGAAATCAAAGTACCTTGGAGAGATTAACAGAGATCTGGACCCCAAGAGGGG 480

Qy 481 AAACATGCTGGACCCCAAGAGTGGAGAGAGAGAGAGATCTGGAGAGAGATCTGAAGAGATC 540  
 Db 481 AAACATGCTGGACCCCAAGAGTGGAGAGAGAGAGAGATCTGGAGAGAGATCTGAAGAGATC 540

Qy 541 AGCGACCTGGAGAGATGACGAGTAACCTCCCTCG 576  
 Db 541 AGCGACCTGGAGAGATGACGAGTAACCTCCCTCG 576

RESULT 6  
 US-09-392-714-17  
 ; Sequence 17, Application US/09392714A  
 ; Patent No. 6686147  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Scanlan, Matthew J.  
 ; APPLICANT: Gure, Ali O.  
 ; APPLICANT: Williamson, Barbara  
 ; APPLICANT: Chen, Yao-Tsung  
 ; APPLICANT: Old, Lloyd J.  
 ; TITLE OF INVENTION: Cancer Associated Antigens and Uses  
 ; FILE REFERENCE: L0461/7062  
 ; CURRENT APPLICATION NUMBER: US/09/392,714A  
 ; EARLIER FILING DATE: 1999-09-09  
 ; EARLIER FILING DATE: 1998-07-15  
 ; NUMBER OF SEQ ID NOS: 30  
 ; SOFTWARE: FastSeq for Windows Version 3.0  
 ; SEQ ID NO 17  
 ; LENGTH: 576  
 ; TYPE: DNA  
 ; ORGANISM: Homo sapiens  
 US-09-392-714-17

Query Match 100.0%; Score 576; DB 4; Length 576;

Best Local Similarity 100.0%; Pred. No. 1.7e-184;  
 Matches 576; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ATGAACGGAGACGACGCTTTGCAAGGAGAGACCCAGGGATGATGCTCAAAATATCAGAGAAG 60  
 Db 1 ATGAACGGAGACGACGCTTTGCAAGGAGAGACCCAGGGATGATGCTCAAAATATCAGAGAAG 60

Qy 61 TTACGAAGGCTTCGATGATATTGCCAAATACCTTCTTAAGAAAGAGTGGAAAGATG 120  
 Db 61 TTACGAAGGCTTCGATGATATTGCCAAATACCTTCTTAAGAAAGAGTGGAAAGATG 120

Qy 121 AAATCCTCGGAGAAATCGTCTATGTATATGAAGCTAAACTATGAGGTCACTATAA 180  
 Db 121 AAATCCTCGGAGAAATCGTCTATGTATATGAAGCTAAACTATGAGGTCACTATAA 180

Qy 181 CTAGGTTTCAAGGTCACTCCACCTTTCATGCGTAGTAAACCGGCTGCAGACTTCCAC 240  
 Db 181 CTAGGTTTCAAGGTCACTCCACCTTTCATGCGTAGTAAACCGGCTGCAGACTTCCAC 240

Qy 241 GGGATGATTTTGGTAACGATCGAAACCAACAGGAATCAGGTTGAACGTCCTCAGTACT 300  
 Db 241 GGGATGATTTTGGTAACGATCGAAACCAACAGGAATCAGGTTGAACGTCCTCAGTACT 300

Qy 301 TTCCGACGCTCCAGAGAAATCTTCCGAAAGATCATGCCAAGAGCCAGCAGAGAGAGAA 360  
 Db 301 TTCCGACGCTCCAGAGAAATCTTCCGAAAGATCATGCCAAGAGCCAGCAGAGAGAGAA 360

Qy 361 AATGGTTGAAGGAGTCCAGAGCATCTCGGCCCAAAATGATGGAAACAGCTGTGC 420  
 Db 361 AATGGTTGAAGGAGTCCAGAGCATCTCGGCCCAAAATGATGGAAACAGCTGTGC 420

Qy 421 CCCCCGGGAAATCAAAGTACCTTGGAGAGATTAACAGAGATCTGGACCCCAAGAGGGG 480  
 Db 421 CCCCCGGGAAATCAAAGTACCTTGGAGAGATTAACAGAGATCTGGACCCCAAGAGGGG 480

Qy 481 AAACATGCTGGACCCCAAGAGTGGAGAGAGAGAGAGATCTGGAGAGAGATCTGAAGAGATC 540  
 Db 481 AAACATGCTGGACCCCAAGAGTGGAGAGAGAGAGAGATCTGGAGAGAGATCTGAAGAGATC 540

Qy 541 AGCGACCTGGAGAGATGACGAGTAACCTCCCTCG 576  
 Db 541 AGCGACCTGGAGAGATGACGAGTAACCTCCCTCG 576

RESULT 7  
 US-09-392-714-16  
 ; Sequence 16, Application US/09392714A  
 ; Patent No. 6686147  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Scanlan, Matthew J.  
 ; APPLICANT: Gure, Ali O.  
 ; APPLICANT: Williamson, Barbara  
 ; APPLICANT: Chen, Yao-Tsung  
 ; APPLICANT: Old, Lloyd J.  
 ; TITLE OF INVENTION: Cancer Associated Antigens and Uses  
 ; FILE REFERENCE: L0461/7062  
 ; CURRENT APPLICATION NUMBER: US/09/392,714A  
 ; EARLIER FILING DATE: 1999-09-09  
 ; EARLIER FILING DATE: 1998-07-15  
 ; NUMBER OF SEQ ID NOS: 30  
 ; SOFTWARE: FastSeq for Windows Version 3.0  
 ; SEQ ID NO 16  
 ; LENGTH: 766  
 ; TYPE: DNA  
 ; ORGANISM: Homo sapiens  
 US-09-392-714-16

Query Match 86.2%; Score 496.6; DB 4; Length 766;  
 Best Local Similarity 91.5%; Pred. No. 1.3e-157;  
 Matches 526; Conservative 0; Mismatches 49; Indels 0; Gaps 0;

```

QY 1 ATGAACGGAGACGAGCGCTTTCAGAGGACCCAGGATGATGCTCAATATCAGAGAAG 60
Db 92 ATGAACGGAGACGAGCGCTTTCAGAGGACCCAGGATGATGCTCAATATCAGAGAAG 151
QY 61 TTACGAAAGGCGCTTCGATGATATTCGCAATATCTTCTTAAGAAAGAGTGGGAAAGATG 120
Db 152 ATCCAAAGGCGCTTCGATGATATTCGCAATATCTTCTTAAGAAAGAGTGGGAAAGATG 211
QY 121 AAATCTCGGAGAAATCTCTATGATATATCAAGCTTAACTATGAGTCTGACTTAA 180
Db 212 AAAGCTTCGAGAAATCTCTATGATATATCAAGCTTAACTATGAGTCTGACTTAA 271
QY 181 CTAGGTTTCAAGTTCACCTCCACCTTTCATGCGTAGTAAACGGGCTCGAGACTTCCAC 240
Db 272 CTAGGTTTCAAGGCTACCTCCACCTTTCATGCGTAGTAAACGGGCTCGAGACTTCCAC 331
QY 241 GGAATGATTTGGTAAAGATGAAACCCACAGAAATCAGGTTGAACTCTCTGATGACT 300
Db 332 GGAATGATTTGGTAAAGATGAAACCCACAGAAATCAGGTTGAACTCTCTGATGACT 391
QY 301 TTCCGAGGCTCCAGAGATCTTCCGAGAGATCATGCCAAGAGCCAGCAGAGGAAGAA 360
Db 392 TTCCGAGGCTCCAGAGATCTTCCGAGAGATCATGCCAAGAGCCAGCAGAGGAAGA 451
QY 361 AATGGTTTGAAGGAGTGGCAGAGGATCTGGCCCAAAATGATGGGAAACAGCTGTGC 420
Db 452 AATGATTCGGAGGAGTGGCAGAGGATCTGGCCCAAAATGATGGGAAAGAGCTGTGC 511
QY 421 CCCCCGGGAAATCCAGTACCTTGGAGAGATTAACAGACATCTGGACCCCAAGGGGG 480
Db 512 CCCCCGGGAAATCCAGTACCTTGGAGAGATTAACAGACATCTGGACCCCAAGGGGG 571
QY 481 AATGATGCTGGAGCCACAGACTGCGTGAGAGAAAGCAGCTGCTGTTTATGAGAGATC 540
Db 572 GAACATGCTGGAGCCACAGACTGCGTGAGAGAAAGCAGCTGCTGTTTATGAGAGATC 631
QY 541 AGCGACCTCGAGGAGATGACGAGTAACTCCCTC 575
Db 632 AGCGACCTCGAGGAGATGACGAGTAACTCCCTC 666

```

RESULT 8

```

US-08-479-328-2
; Sequence 2, Application US/08479328
; Patent No. 5698396
; GENERAL INFORMATION:
; APPLICANT: Pfeundschuh, Michael
; TITLE OF INVENTION: Method For Identifying Or Isolating A Molecule
; TITLE OF INVENTION: And Molecules Identified Thereby
; NUMBER OF SEQUENCES: 3
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Felfe & Lynch
; STREET: 805 Third Avenue
; CITY: New York City
; STATE: New York
; ZIP: 10022
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.5 inch, 360 kb storage
; COMPUTER: IBM
; OPERATING SYSTEM: PC-DOS
; SOFTWARE: Wordperfect
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/479,328
; FILING DATE: 07-JUNE-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Hanson, No. 5698396man D.
; REGISTRATION NUMBER: 30,946
; REFERENCE/DOCKET NUMBER: LUD 5410
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 688-9200
; TELEFAX: (212) 838-3884
; INFORMATION FOR SEQ ID NO: 2:

```

```

; SEQUENCE CHARACTERISTICS:
; LENGTH: 931 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double-stranded
; TOPOLOGY: linear
; US-08-479-328-2

Query Match      86.2%; Score 496.6; DB 1; Length 931;
Best Local Similarity 91.5%; Pred. No. 1.5e-157;
Matches 526; Conservative 0; Mismatches 49; Indels 0; Gaps 0;

QY 1 ATGAACGGAGACGAGCGCTTTCAGAGGACCCAGGATGATGCTCAATATCAGAGAAG 60
Db 99 ATGAACGGAGACGAGCGCTTTCAGAGGACCCAGGATGATGCTCAATATCAGAGAAG 158
QY 61 TTACGAAAGGCGCTTCGATGATATTCGCAATATCTTCTTAAGAAAGAGTGGGAAAGATG 120
Db 159 ATCCAAAGGCGCTTCGATGATATTCGCAATATCTTCTTAAGAAAGAGTGGGAAAGATG 218
QY 121 AAATCTCGGAGAAATCTCTATGATATATCAAGCTTAACTATGAGTCTGACTTAA 180
Db 219 AAAGCTTCGAGAAATCTCTATGATATATCAAGCTTAACTATGAGTCTGACTTAA 278
QY 181 CTAGGTTTCAAGTTCACCTCCACCTTTCATGCGTAGTAAACGGGCTCGAGACTTCCAC 240
Db 279 CTAGGTTTCAAGGCTACCTCCACCTTTCATGCGTAGTAAACGGGCTCGAGACTTCCAC 338
QY 241 GGAATGATTTGGTAAAGATGAAACCCACAGAAATCAGGTTGAACTCTCTGATGACT 300
Db 339 GGAATGATTTGGTAAAGATGAAACCCACAGAAATCAGGTTGAACTCTCTGATGACT 398
QY 301 TTCCGAGGCTCCAGAGATCTTCCGAGAGATCATGCCAAGAGCCAGCAGAGGAAGAA 360
Db 399 TTCCGAGGCTCCAGGAGATCTTCCGAGAGATCATGCCAAGAGCCAGCAGAGGAAGA 458
QY 361 AATGGTTTGAAGGAGTGGCAGAGGATCTGGCCCAAAATGATGGGAAACAGCTGTGC 420
Db 459 AATGATTCGGAGGAGTGGCAGAGGATCTGGCCCAAAATGATGGGAAAGAGCTGTGC 518
QY 421 CCCCCGGGAAATCCAGTACCTTGGAGAGATTAACAGACATCTGGACCCCAAGGGGG 480
Db 519 CCCCCGGGAAATCCAGTACCTTGGAGAGATTAACAGACATCTGGACCCCAAGGGGG 578
QY 481 AAACATGCTGGAGCCACAGACTGCGTGAGAGAAAGCAGCTGCTGTTTATGAGAGATC 540
Db 579 GAACATGCTGGAGCCACAGACTGCGTGAGAGAAAGCAGCTGCTGTTTATGAGAGATC 638
QY 541 AGCGACCTCGAGGAGATGACGAGTAACTCCCTC 575
Db 639 AGCGACCTCGAGGAGATGACGAGTAACTCCCTC 673

```

RESULT 9

```

US-08-761-119-2
; Sequence 2, Application US/08761119
; Patent No. 5798264
; GENERAL INFORMATION:
; APPLICANT: Pfeundschuh, Michael
; TITLE OF INVENTION: Method For Identifying Or Isolating A Molecule
; TITLE OF INVENTION: And Molecules Identified Thereby
; NUMBER OF SEQUENCES: 3
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Felfe & Lynch
; STREET: 805 Third Avenue
; CITY: New York City
; STATE: New York
; ZIP: 10022
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.5 inch, 360 kb storage
; COMPUTER: IBM
; OPERATING SYSTEM: PC-DOS
; SOFTWARE: Wordperfect
; CURRENT APPLICATION DATA:

```



```

; APPLICATION NUMBER: US/08/761,119
; FILING DATE: 6-DECEMBER-1996
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/479,328
; FILING DATE: 07-JUNE-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Hanson, No. 5798264man D.
; REGISTRATION NUMBER: 30,946
; REFERENCE/DOCKET NUMBER: LUD 5410.3
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 688-9200
; TELEFAX: (212) 838-3884
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 931 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double-stranded
; TOPOLOGY: linear
; US-08-761-119-2

```

```

Query Match      86.2%; Score 496.6; DB 1; Length 931;
Best Local Similarity 91.5%; Pred. No. 1.5e-157;
Matches 526; Conservative 0; Mismatches 49; Indels 0; Gaps 0;

QY 1 ATGAACGGAGACGACGCTTTGCAAGGAGACCCAGGATGATGCTCAAAATATCAGAGAAG 60
DB 99 ATGAACGGAGACGACGCTTTGCAAGGAGACCCAGGATGATGCTCAAAATATCAGAGAAG 158
QY 61 TTACGAAGGCTTCGATGATATGCAAAATCTTCTAAGAAAGATGGGAAGATG 120
DB 159 ATCAAAAGGCTTCGATGATATGCAAAATCTTCTAAGAAAGATGGGAAGATG 218
QY 121 AAATCTCGAGAAATCGTCTATGTATATGAAGCTAAACTATGAGGTGATGACTAAA 180
DB 219 AAAGCTTCGAGAAATCTTCTAAGAAAGATGATGAGGTGATGACTAAA 278
QY 181 CTAGGTTTCAAGTTCACCTCCACCTTTCATGCGTAGTAAACGGGTGAGACTTCCAC 240
DB 279 CTAGGTTTCAAGTTCACCTCCACCTTTCATGCGTAGTAAACGGGTGAGACTTCCAC 338
QY 241 GGGATGATTTTGGTACGATCGAAACACAGGAATCGAGGTGAAACGCTCCAGATGACT 300
DB 339 GGGATGATTTTGGTACGATCGAAACACAGGAATCGAGGTGAAACGCTCCAGATGACT 398
QY 301 TTCCGACGCTCCAGAGATCTTCCGAGATCATGCCCAAGAGCCAGCAGAGGAAGAA 360
DB 399 TTCCGACGCTCCAGAGATCTTCCGAGATCATGCCCAAGAGCCAGCAGAGGAAGAA 458
QY 361 AATGTTTGAAGGAGTTCAGAGGATCTGGCCCAACAAATGATGGGAACAGCTGTGC 420
DB 459 AATGTTTGAAGGAGTTCAGAGGATCTGGCCCAACAAATGATGGGAACAGCTGTGC 518
QY 421 CCCCCGGAATCCAAATGACCTTCGAGAAAGATTAACAGAGATCTGGACCCCAAGGGGG 480
DB 519 CCCCCGGAATCCAAATGACCTTCGAGAAAGATTAACAGAGATCTGGACCCCAAGGGGG 578
QY 481 AAACATGCTCGACCCCAAGAGTTCGAGAGAAAGAGCTGGTGGTTATGAGAGATC 540
DB 579 GAACATGCTCGACCCCAAGAGTTCGAGAGAAAGAGCTGGTGGTTATGAGAGATC 638
QY 541 AGCGACCTTGAGGAAGATGACGAGTAACCTCCCTC 575
DB 639 AGCGACCTTGAGGAAGATGACGAGTAACCTCCCTC 673

```

```

RESULT 10
US-08-668-128B-2
; Sequence 2, Application US/08668128B
; Patent No. 5840568
; GENERAL INFORMATION:
; APPLICANT: Pfreundschuh, Michael
; TITLE OF INVENTION: Hodgkin's Disease Associated Molecules And

```

```

; TITLE OF INVENTION: Uses Thereof
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Felfe & Lynch
; STREET: 805 Third Avenue
; CITY: New York City
; STATE: New York
; ZIP: 10022
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.5 inch, 360 kb storage
; COMPUTER: IBM
; OPERATING SYSTEM: PC-DOS
; SOFTWARE: Wordperfect
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/668,128B
; FILING DATE: 21-JUNE-1996
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/644,116
; FILING DATE: 10-MAY-1996
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/580,980
; FILING DATE: 03-JANUARY-1996
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/479,328
; FILING DATE: 07-JUNE-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Hanson, No. 5840568man D.
; REGISTRATION NUMBER: 30,946
; REFERENCE/DOCKET NUMBER: LUD 5441
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 688-9200
; TELEFAX: (212) 838-3884
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 931 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double-stranded
; TOPOLOGY: linear
; US-08-668-128B-2

```

```

Query Match      86.2%; Score 496.6; DB 2; Length 931;
Best Local Similarity 91.5%; Pred. No. 1.5e-157;
Matches 526; Conservative 0; Mismatches 49; Indels 0; Gaps 0;

QY 1 ATGAACGGAGACGACGCTTTGCAAGGAGACCCAGGATGATGCTCAAAATATCAGAGAAG 60
DB 99 ATGAACGGAGACGACGCTTTGCAAGGAGACCCAGGATGATGCTCAAAATATCAGAGAAG 158
QY 61 TTACGAAGGCTTCGATGATATGCAAAATCTTCTAAGAAAGATGGGAAGATG 120
DB 159 ATCAAAAGGCTTCGATGATATGCAAAATCTTCTAAGAAAGATGGGAAGATG 218
QY 121 AAATCTCGAGAAATCGTCTATGTATATGAAGCTAAACTATGAGGTGATGACTAAA 180
DB 219 AAAGCTTCGAGAAATCTTCTAAGAAAGATGATGAGGTGATGACTAAA 278
QY 181 CTAGGTTTCAAGTTCACCTCCACCTTTCATGCGTAGTAAACGGGTGAGACTTCCAC 240
DB 279 CTAGGTTTCAAGTTCACCTCCACCTTTCATGCGTAGTAAACGGGTGAGACTTCCAC 338
QY 241 GGGATGATTTTGGTACGATCGAAACACAGGAATCGAGGTGAAACGCTCCAGATGACT 300
DB 339 GGGATGATTTTGGTACGATCGAAACACAGGAATCGAGGTGAAACGCTCCAGATGACT 398
QY 301 TTCCGACGCTCCAGAGATCTTCCGAGATCATGCCCAAGAGCCAGCAGAGGAAGAA 360
DB 399 TTCCGACGCTCCAGAGATCTTCCGAGATCATGCCCAAGAGCCAGCAGAGGAAGAA 458
QY 361 AATGTTTGAAGGAGTTCAGAGGATCTGGCCCAACAAATGATGGGAACAGCTGTGC 420
DB 459 AATGTTTGAAGGAGTTCAGAGGATCTGGCCCAACAAATGATGGGAACAGCTGTGC 518

```







; TITLE OF INVENTION: Method For Identifying Or Isolating A Molecule
; TITLE OF INVENTION: And Molecules Identified Thereby
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Felfe & Lynch
; STREET: 805 Third Avenue
; CITY: New York City
; STATE: New York
; ZIP: 10022
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.5 inch, 360 kb storage
; COMPUTER: IBM
; OPERATING SYSTEM: PC-DOS
; SOFTWARE: Wordperfect
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/580,980A
; FILING DATE: 03-JANUARY-1996
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/479,328
; FILING DATE: 07-JUNE-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Hanson, No. 602519man D.
; REGISTRATION NUMBER: 30,946
; REFERENCE/DOCKET NUMBER: LUD 5410.1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 688-9200
; TELEFAX: (212) 838-3884
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 931 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double-stranded
; TOPOLOGY: linear
; US-08-580-980A-2

Query Match 86.2%; Score 496.6; DB 3; Length 931;
Best Local Similarity 91.5%; Pred. No. 1.5e-157;
Matches 526; Conservative 0; Mismatches 49; Indels 0; Gaps 0;

QY 1 ATGAACGAGACGCGCTTTGCAAGGAGAGACCCAGGGATGATGCTCAAAATATCAGAGAAG 60
Db 99 ATGAACGAGACGCGCTTTGCAAGGAGAGACCCAGGGATGATGCTCAAAATATCAGAGAAG 158
QY 61 TTACGAAGGCGCTTCGATGATATTCGCAAAATCTCTTACGAAGAGTGCGGAAAGATG 120
Db 159 ATCCAAAAGGCGCTTCGATGATATTCGCAAAATCTCTTACGAAGAGTGCGGAAAGATG 218
QY 121 AAATCTCGGAGAAATCGTCTATGTATATGAAGCTAAACTATGAGCTCATGACTAAA 180
Db 219 AAAGCGCTCGGAGAAATCTCTTATGTATATGAGAGAAAGTATGAGGCTATGACTAAA 278
QY 181 CTAGGTTTCAAGGTCACCCCTCCACCTTTTCATGCTAGTAAACCGGGCTGCAGACTTCCAC 240
Db 279 CTAGGTTTCAAGGTCACCCCTCCACCTTTTCATGCTAGTAAACCGGGCTGCAGACTTCCAC 338
QY 241 GGAATGATTTTGGTAAACGATCGAATCAGGATGAGGATGAGGATGAGGATGAGGATG 300

; TITLE OF INVENTION: Method For Identifying Or Isolating A Molecule
; TITLE OF INVENTION: And Molecules Identified Thereby
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Felfe & Lynch
; STREET: 805 Third Avenue
; CITY: New York City
; STATE: New York
; ZIP: 10022
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.5 inch, 360 kb storage
; COMPUTER: IBM
; OPERATING SYSTEM: PC-DOS
; SOFTWARE: Wordperfect
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/580,980A
; FILING DATE: 03-JANUARY-1996
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/479,328
; FILING DATE: 07-JUNE-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Hanson, No. 602519man D.
; REGISTRATION NUMBER: 30,946
; REFERENCE/DOCKET NUMBER: LUD 5410.1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 688-9200
; TELEFAX: (212) 838-3884
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 931 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double-stranded
; TOPOLOGY: linear
; US-08-580-980A-2

Query Match 86.2%; Score 496.6; DB 3; Length 931;
Best Local Similarity 91.5%; Pred. No. 1.5e-157;
Matches 526; Conservative 0; Mismatches 49; Indels 0; Gaps 0;

QY 1 ATGAACGAGACGCGCTTTGCAAGGAGAGACCCAGGGATGATGCTCAAAATATCAGAGAAG 60
Db 99 ATGAACGAGACGCGCTTTGCAAGGAGAGACCCAGGGATGATGCTCAAAATATCAGAGAAG 158
QY 61 TTACGAAGGCGCTTCGATGATATTCGCAAAATCTCTTACGAAGAGTGCGGAAAGATG 120
Db 159 ATCCAAAAGGCGCTTCGATGATATTCGCAAAATCTCTTACGAAGAGTGCGGAAAGATG 218
QY 121 AAATCTCGGAGAAATCGTCTATGTATATGAAGCTAAACTATGAGCTCATGACTAAA 180
Db 219 AAAGCGCTCGGAGAAATCTCTTATGTATATGAGAGAAAGTATGAGGCTATGACTAAA 278
QY 181 CTAGGTTTCAAGGTCACCCCTCCACCTTTTCATGCTAGTAAACCGGGCTGCAGACTTCCAC 240
Db 279 CTAGGTTTCAAGGTCACCCCTCCACCTTTTCATGCTAGTAAACCGGGCTGCAGACTTCCAC 338
QY 241 GGAATGATTTTGGTAAACGATCGAATCAGGATGAGGATGAGGATGAGGATGAGGATG 300
Db 339 GGAATGATTTTGGTAAACGATCGAATCAGGATGAGGATGAGGATGAGGATGAGGATG 398
QY 301 TTCCGAGCGCTCCAGAGATCTCCGAGAGATCATGCCCAAGAGCCAGCAGAGAGAGAA 360
Db 399 TTCCGAGCGCTCCAGAGATCTCCGAGAGATCATGCCCAAGAGCCAGCAGAGAGAGAA 458
QY 361 AATGTTTGAAGGAAGTGCAGAGGATCTGCGCCACAAAATATGATGGGAAAACAGCTGTGC 420
Db 459 AATGTTTGAAGGAAGTGCAGAGGATCTGCGCCACAAAATATGATGGGAAAACAGCTGTGC 518
QY 421 CCCCCGGGAAATCCAAAGTACCTTGGGAGAAATTAACAAGATCTGAGCCCAAAAGGGGG 480
Db 519 CCCCCGGGAAATCCAAAGTACCTTGGGAGAAATTAACAAGATCTGAGCCCAAAAGGGGG 578
QY 481 ABAATGCTGGACCCACAGACTCGCTGAGAGAAAGCAGCTGTGTTTATGAAGAGATC 540

339	GGGAATGATTTTGGATAATCAACCTAACCGTGGGAATCAGGTTGAACGTCTCAGATGACT	398
301	TTCCGCAGCCTCCAGAGAAATCTTCCCGGAAGATCATGCCCAAGAACCGACGAGAGAAGAA	360
399	TTCCGCAGGCTCCAGGGAAATCTCCCGGAAGATCATGCCCAAGAACCGACGAGAGAAGA	458
361	AATGTTTTGAAGGAAGTGCCAGAGGCATCTGCCCCACAAAAATGATGGGAAACAGCTGTGC	420
459	AATGATTCGGAGGAAGTCCAGAAGCATCTGGCCCCACAAAATGATGGGAAAGAGCTGTGC	518
421	CCCCCGGAAATCCAAGTACCTTTGAGAGAGATTAAACAAGACATCTGGACCCCAAAAGGGGG	480
519	CCCCCGGAAACCAACTACCTCTGAGAAGATTCAGAGAGATCTGGAGCCCAAAAGGGGG	578
481	AAACATGCTCGACCCACAGACTGGCTGAGAGAAAGCAGCTGGTGGTTATTGAAGAGATC	540
579	GAAATGCTTGGACCCACAGACTGGCTGAGAGAAACAGCTGGTGCATTTATGAAGAGATC	638
541	AGCAGCCTTGAGGAAGATGACGAGTAACTCCGCTC	575
639	AGCAGCCTTGAGGAAGATGACGAGTAACTCCGCTC	673

Search completed: March 31, 2004, 17:26:48  
Job time : 99 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 31, 2004, 16:39:19 ; Search time 389 Seconds  
(without alignments)

5512.734 Million cell updates/sec

Title: US-09-975-856-1

Perfect score: 576

Sequence: 1 ATGACGAGACGACGCTT.....ATGACGAGTAACTCCCTCG 576

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 2458946 seqs, 1861504846 residues

Total number of hits satisfying chosen parameters: 4917892

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications NA:\*

```

1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:*
2: /cgn2_6/ptodata/1/pubpna/PCT_NEW_PUB.seq:*
3: /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq:*
4: /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq:*
5: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq:*
6: /cgn2_6/ptodata/1/pubpna/PCTUS_PUBCOMB.seq:*
7: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq:*
8: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq:*
9: /cgn2_6/ptodata/1/pubpna/US09A_PUBCOMB.seq:*
10: /cgn2_6/ptodata/1/pubpna/US09B_PUBCOMB.seq:*
11: /cgn2_6/ptodata/1/pubpna/US09C_PUBCOMB.seq:*
12: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq:*
13: /cgn2_6/ptodata/1/pubpna/US10A_PUBCOMB.seq:*
14: /cgn2_6/ptodata/1/pubpna/US10B_PUBCOMB.seq:*
15: /cgn2_6/ptodata/1/pubpna/US10C_PUBCOMB.seq:*
16: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq:*
17: /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq:*
18: /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq:*

```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	576	100.0	576	10	US-09-975-856-1
2	576	100.0	576	10	US-09-833-039-5
3	576	100.0	576	14	US-10-177-277-5
4	576	100.0	576	15	US-10-117-937-599
5	496.6	86.2	766	10	US-09-849-602-13
6	496.6	86.2	766	15	US-10-207-655-84
7	496.6	86.2	766	15	US-10-117-937-6
8	496.6	86.2	931	10	US-09-833-039-2
9	496.6	86.2	931	14	US-10-177-277-2
10	496.6	86.2	1309	9	US-09-954-531-1000
11	496.6	86.2	1309	9	US-09-954-531-1392
12	496	86.1	1249	15	US-10-159-563-402
13	482.2	83.7	576	10	US-09-975-856-2
14	482.2	83.7	576	10	US-09-833-039-6
15	482.2	83.7	576	14	US-10-177-277-6

16	480	83.3	766	10	US-09-833-039-1
17	480	83.3	766	14	US-10-177-277-1
18	308.6	53.6	574	15	US-10-027-632-106782
19	308.6	53.6	718	15	US-10-027-632-24594
20	140	24.3	564	14	US-10-029-386-10234
21	137	23.8	189	14	US-10-029-386-23851
22	111.4	19.3	675	15	US-10-027-632-33096
23	111.4	19.3	675	15	US-10-027-632-33097
24	108.6	18.9	734	15	US-10-027-632-148783
25	87.8	15.2	3186778	15	US-10-027-632-174961
26	72	12.5	3225	15	US-10-027-632-114514
27	72	12.5	3225	15	US-10-027-632-114515
28	47	8.2	575	15	US-10-027-632-32431
29	47	8.2	575	15	US-10-027-632-32432
30	41	7.1	583	10	US-09-814-353-18480
31	40.8	7.1	60	10	US-09-908-975-6728
32	39.8	6.9	3591	9	US-09-981-353-84
33	39.6	6.9	439	12	US-10-085-783A-58497
34	39.6	6.9	439	15	US-10-242-535A-58497
35	39.4	6.8	1097	14	US-10-133-013-261
36	39	6.8	629	15	US-10-027-632-134770
37	37.8	6.6	389	12	US-10-085-783A-18965
38	37.8	6.6	389	15	US-10-242-535A-18965
39	37.8	6.6	538	14	US-10-029-386-7375
40	37.8	6.6	1530	10	US-09-814-353-21781
41	37.8	6.6	2024	15	US-10-108-260A-1253
42	37.8	6.6	2736	13	US-10-098-841-181
43	37.8	6.6	3205	14	US-10-269-909-87
44	37.8	6.6	5532	10	US-09-971-392-98
45	37.6	6.5	856	15	US-10-027-632-163719

## ALIGNMENTS

## RESULT 1

US-09-975-856-1  
; Sequence 1, Application US/09975856  
; Publication No. US20030023057A1

GENERAL INFORMATION:

APPLICANT: Gure, Ali O.; Tureci, Ozlem; Sahin, Ugur; Tsang, Solam;

Scanlan, Matthew J.; Knuth Alexander; Pfreundschuh, Michael; Old;

Lloyd J.; Chen, Yao-Tseng

TITLE OF INVENTION: Isolated Nucleic Acid Molecules Encoding SSX

Family

Members And Uses Thereof

NUMBER OF SEQUENCES: 14

CORRESPONDENCE ADDRESS:

ADDRESSER: Fulbright & Jaworski L.L.P.

STREET: 666 Fifth Avenue

CITY: New York City

STATE: New York

ZIP: 10103

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette, 3.5 inch, 144 kb storage

COMPUTER: IBM

OPERATING SYSTEM: PC-DOS

SOFTWARE: Wordperfect

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/975,856

FILING DATE: 11-Oct-2001

CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 09/796,780

FILING DATE: 2001-03-01

ATTORNEY/AGENT INFORMATION:

NAME: Hanson, No. US20030023057Alman D.

REGISTRATION NUMBER: 30,946

REFERENCE/DOCKET NUMBER: LUD 5480

TELECOMMUNICATION INFORMATION:

TELEPHONE: (212) 318-3168

TELEFAX: (212) 752-5958

INFORMATION FOR SEQ ID NO: 1:

```
/
/ SEQUENCE CHARACTERISTICS:
/   LENGTH: 576 nucleotides
/   TYPE: nucleic acid
/   STRANDEDNESS: single
/   TOPOLOGY: linear
/
/   SEQUENCE DESCRIPTION: SEQ ID NO: 1:
US-09-975-856-1

Query Match      100.0%; Score 576; DB 10; Length 576;
Best Local Similarity 100.0%; Pred. No. 2.8e-174;
Matches 576; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGAACGAGAGACGAGCCCTTTCGACGAGAGACCCAGGATGATGCTCAAAATATCAGAGAAG 60
Db 1 ATGAACGAGAGACGAGCCCTTTCGACGAGAGACCCAGGATGATGCTCAAAATATCAGAGAAG 60
QY 61 TTACGAAAGGCCCTTCGATGATATTCGCAAAATACCTTCTTAAGAAAGAGTGGGAAAGATG 120
Db 61 TTACGAAAGGCCCTTCGATGATATTCGCAAAATACCTTCTTAAGAAAGAGTGGGAAAGATG 120
QY 121 AAATCCTCGAGAAATCGTCTATGTATATGAAGCTAAACTATGAGTCACTGACTAAA 180
Db 121 AAATCCTCGAGAAATCGTCTATGTATATGAAGCTAAACTATGAGTCACTGACTAAA 180
QY 181 CTAGGTTTCAAGGTCAACCTCCACCTTTCATGCGTAGTAAACGAGTCACTGACTTCCAC 240
Db 181 CTAGGTTTCAAGGTCAACCTCCACCTTTCATGCGTAGTAAACGAGTCACTGACTTCCAC 240
QY 241 GGGAAATGATTTTGGTAAAGATCGAAACCAAGGATCAGGTTGAACGTCCTCAGATGACT 300
Db 241 GGGAAATGATTTTGGTAAAGATCGAAACCAAGGATCAGGTTGAACGTCCTCAGATGACT 300
QY 301 TTCCGAGCCTCCAGAGATCTTCCGAGATCATGCGCAAGATCATGCCAAGACCCAGAGAGAA 360
Db 301 TTCCGAGCCTCCAGAGATCTTCCGAGATCATGCGCAAGATCATGCCAAGACCCAGAGAGAA 360
QY 361 AATGTTTGAAGAGTGCAGAGGATCTGCGCCCAAGATCATGCGCAAGATCATGCCAAGAG 420
Db 361 AATGTTTGAAGAGTGCAGAGGATCTGCGCCCAAGATCATGCGCAAGATCATGCCAAGAG 420
QY 421 CCCCCGGGAATCAAGTACCTTGGAGAGATTAACAGACATCTGGACCCCAAGAGGGG 480
Db 421 CCCCCGGGAATCAAGTACCTTGGAGAGATTAACAGACATCTGGACCCCAAGAGGGG 480
QY 481 AAACATGCTGGACCCAGACATCTGCGCAAGATCATGCGCAAGATCATGCCAAGAGATC 540
Db 481 AAACATGCTGGACCCAGACATCTGCGCAAGATCATGCGCAAGATCATGCCAAGAGATC 540
QY 541 AGCGACCTCGAGGAGATGACGATTAACCTCCCTCG 576
Db 541 AGCGACCTCGAGGAGATGACGATTAACCTCCCTCG 576
```

## RESULT 2

```
US-09-833-039-5
/ Sequence 5, Application US/09833039
/ Publication No. US20030175960A1
/ GENERAL INFORMATION:
/   APPLICANT: Tureci, Ozlem
/   APPLICANT: Sahin, Ugur
/   APPLICANT: Pfreundschuh, Michael
/   TITLE OF INVENTION: Tumor Associated Peptide and Uses Thereof
/   FILE REFERENCE: LUD 5622.1
/   CURRENT APPLICATION NUMBER: US/09/833,039
/   PRIOR FILING DATE: 2001-04-12
/   PRIOR APPLICATION NUMBER: US 09/409,455
/   PRIOR FILING DATE: 1999-09-30
/   PRIOR APPLICATION NUMBER: US 09/344,040
/   PRIOR FILING DATE: 1999-06-25
/   PRIOR APPLICATION NUMBER: US 09/105,839
/   PRIOR FILING DATE: 1998-06-26
/   PRIOR APPLICATION NUMBER: US 08/851,130
/   PRIOR FILING DATE: 1997-05-05
```

## RESULT 3

```
US-10-177-277-5
/ Sequence 5, Application US/10177277
/ Publication No. US20030185844A1
/ GENERAL INFORMATION:
/   APPLICANT: Tureci, Ozlem
/   APPLICANT: Sahin, Ugur
/   APPLICANT: Pfreundschuh, Michael
/   APPLICANT: Ramersee, Hans Georg
/   APPLICANT: Stevanovic, Stefan
/   TITLE OF INVENTION: Method for Determining Presence of Cancer in a Sample By Determin:
/   TITLE OF INVENTION: Expression of an SSX Gene, Peptides Derived From Said SSX Gene
/   FILE REFERENCE: LUD 5556.1
/   CURRENT APPLICATION NUMBER: US/10/177,277
/   PRIOR FILING DATE: 2002-06-21
/   PRIOR APPLICATION NUMBER: US/09/344,040
/   PRIOR FILING DATE: 1999-06-25
/   PRIOR APPLICATION NUMBER: US 09/105,839
/   PRIOR FILING DATE: 1998-06-26
/   PRIOR APPLICATION NUMBER: US 08/851,130
```

```

RES001 5
US-09-849-602-13
; Sequence 13, Application US/09849602
; Publication No. US20030165834A1
; GENERAL INFORMATION:
; APPLICANT: Scanlan, Matthew J.
; APPLICANT: Old, Lloyd J.
; APPLICANT: Stockert, Elisabeth
; APPLICANT: Chen, Yao-Tseeng
; TITLE OF INVENTION: Colon Cancer An
; FILE REFERENCE: L0461/7105(JRV)
; CURRENT APPLICATION NUMBER: US/09/84
; CURRENT FILING DATE: 2001-05-04
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 13
; LENGTH: 766
; TYPE: DNA
; ORGANISM: Homo sapien
US-09-849-602-13

```





QY 61 TTACGAAGGCGCTTCGATGATATTGCCAAATACCTCTTAAGAAAGAGTGGGAAGATG 120  
 Db 152 ATCCAAAGGCGCTTCGATGATATTGCCAAATACCTCTTAAGAAAGAGTGGGAAGATG 211  
 QY 121 AAATCCTCGGAGAAATCGTCTATGTGTATATGAAGCTAAACTATGAGGTCTATGACTAAA 180  
 Db 212 AAAGCCTCGGAGAAATCTTCTATGTGTATATGAAGAGAAAGTATGAGGCTATGACTAAA 271  
 QY 181 CTAGGTTTCAAGGTCACCCCTCCACCTTTTCATGCGTGTAGTAAACGGGCTCCAGACTTCCAC 240  
 Db 272 CTAGGTTTCAAGGTCACCCCTCCACCTTTTCATGCGTGTAGTAAACGGGCTCCAGACTTCCAG 331  
 QY 241 GGAATGATTTGGTAAAGATCGAAACCAACAGGAATCAGGTTTGAAGCTCTCAGATGACT 300  
 Db 332 GGAATGATTTGGTAAAGATCGAAACCAACAGGAATCAGGTTTGAAGCTCTCAGATGACT 391  
 QY 301 TTCGCGAGCTCCAGAGATCTTCCGAGAGATCATGCCCAAGAGCCAGAGAGAGAA 360  
 Db 392 TTCGCGAGCTCCAGAGATCTTCCGAGAGATCATGCCCAAGAGCCAGAGAGAGAA 451  
 QY 361 AATGTTTGAAGGAGTGCAGAGGATCTGGCCCAACAAATGATGGGAAACAGCTGTGC 420  
 Db 452 AATGATTCGAGAGAGTGCAGAGATCTGGCCCAACAAATGATGGGAAACAGCTGTGC 511  
 QY 421 CCCCCGGGAAATCCAAATACCTTTGAGAGAGATTAACAAGACATCTGGACCCAAAAGGGGG 480  
 Db 512 CCCCCGGGAAATCCAAATACCTTTGAGAGAGATTAACAAGACATCTGGACCCAAAAGGGGG 571  
 QY 481 AAACATGCTGACCCAGAGATCTGGTGGAGAGAGAGTAAACAAGAGATCTGGTGGTATGAAGATC 540  
 Db 572 GAACATGCTGACCCAGAGATCTGGTGGAGAGAGAGTAAACAAGAGATCTGGTGGTATGAAGATC 631  
 QY 541 AGCGACCTTGAGGAAGATGACGAGTAACTCCCTC 575  
 Db 632 AGCGACCTTGAGGAAGATGACGAGTAACTCCCTC 666

RESULT 8

US-09-833-039-2  
 ; Sequence 2, Application US/09833039  
 ; Publication No. US20030175960A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Tureci, Ozlem  
 ; APPLICANT: Sahin, Ugur  
 ; APPLICANT: Pfreundschuh, Michael  
 ; TITLE OF INVENTION: Tumor Associated Peptide and Uses Thereof  
 ; FILE REFERENCE: LUD 5622.1  
 ; CURRENT APPLICATION NUMBER: US/09/833,039  
 ; CURRENT FILING DATE: 2001-04-12  
 ; PRIOR APPLICATION NUMBER: US 09/409,455  
 ; PRIOR FILING DATE: 1999-09-30  
 ; PRIOR APPLICATION NUMBER: US 09/344,040  
 ; PRIOR FILING DATE: 1999-06-25  
 ; PRIOR APPLICATION NUMBER: US 09/105,839  
 ; PRIOR FILING DATE: 1998-06-26  
 ; PRIOR APPLICATION NUMBER: US 08/851,130  
 ; PRIOR FILING DATE: 1997-05-05  
 ; NUMBER OF SEQ ID NOS: 129  
 ; SEQ ID NO 2  
 ; LENGTH: 931  
 ; TYPE: DNA  
 ; ORGANISM: Homo sapiens  
 US-09-833-039-2

Query Match 86.2%; Score 496.6; DB 10; Length 931;  
 Best Local Similarity 91.5%; Pred. No. 1.1e-148;  
 Matches 526; Conservative 0; Mismatches 49; Indels 0; Gaps 0;  
 QY 1 ATGAACGAGAGCGCTTTGCAAGGAGACCCAGGAGTATGCTCAAAATATCAGAGAAG 60  
 Db 99 ATGAACGAGAGCGCTTTGCAAGGAGACCCAGGAGTATGCTCAAAATATCAGAGAAG 158  
 QY 61 TTACGAAGGCGCTTCGATGATATTGCCAAATACCTCTTAAGAAAGAGTGGGAAGATG 120

Db 159 ATCCAAAGGCGCTTCGATGATATTGCCAAATACCTCTTAAGAAAGAGTGGGAAGATG 218  
 QY 121 AAATCCTCGGAGAAATCGTCTATGTGTATATGAAGCTAAACTATGAGGTCTATGACTAAA 180  
 Db 219 AAAGCCTCGGAGAAATCTTCTATGTGTATATGAAGAGAAAGTATGAGGCTATGACTAAA 278  
 QY 181 CTAGGTTTCAAGGTCACCCCTCCACCTTTTCATGCGTGTAGTAAACGGGCTCCAGACTTCCAC 240  
 Db 279 CTAGGTTTCAAGGTCACCCCTCCACCTTTTCATGCGTGTAGTAAACGGGCTCCAGACTTCCAG 338  
 QY 241 GGAATGATTTGGTAAAGATCGAAACCAACAGGAATCAGGTTTGAAGCTCTCAGATGACT 300  
 Db 339 GGAATGATTTGGTAAAGATCGAAACCAACAGGAATCAGGTTTGAAGCTCTCAGATGACT 398  
 QY 301 TTCGCGAGCTCCAGAGATCTTCCGAGAGATCATGCCCAAGAGCCAGAGAGAGAA 360  
 Db 399 TTCGCGAGCTCCAGAGATCTTCCGAGAGATCATGCCCAAGAGCCAGAGAGAGAA 458  
 QY 361 AATGTTTGAAGGAGTGCAGAGGATCTGGCCCAACAAATGATGGGAAACAGCTGTGC 420  
 Db 459 AATGATTCGAGAGAGTGCAGAGATCTGGCCCAACAAATGATGGGAAACAGCTGTGC 518  
 QY 421 CCCCCGGGAAATCCAAATACCTTTGAGAGAGATTAACAAGACATCTGGACCCAAAAGGGGG 480  
 Db 519 CCCCCGGGAAATCCAAATACCTTTGAGAGAGATTAACAAGACATCTGGACCCAAAAGGGGG 578  
 QY 481 AAACATGCTGACCCAGAGATCTGGTGGAGAGAGAGTAAACAAGAGATCTGGTGGTATGAAGATC 540  
 Db 579 GAACATGCTGACCCAGAGATCTGGTGGAGAGAGAGTAAACAAGAGATCTGGTGGTATGAAGATC 638  
 QY 541 AGCGACCTTGAGGAAGATGACGAGTAACTCCCTC 575  
 Db 639 AGCGACCTTGAGGAAGATGACGAGTAACTCCCTC 673

RESULT 9

US-10-177-277-2  
 ; Sequence 2, Application US/10177277  
 ; Publication No. US20030185844A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Tureci, Ozlem  
 ; APPLICANT: Sahin, Ugur  
 ; APPLICANT: Pfreundschuh, Michael  
 ; APPLICANT: Ramensee, Hans Georg  
 ; APPLICANT: Stvanovic, Stefan  
 ; TITLE OF INVENTION: Method for Determining Presence of Cancer In a Sample By Determini  
 ; TITLE OF INVENTION: Expression of an SSX Gene, Peptides Derived From Said SSX Gene  
 ; FILE REFERENCE: LUD 5556.1  
 ; CURRENT APPLICATION NUMBER: US/10/177,277  
 ; CURRENT FILING DATE: 2002-06-21  
 ; PRIOR APPLICATION NUMBER: US/09/344,040  
 ; PRIOR FILING DATE: 1999-06-25  
 ; PRIOR APPLICATION NUMBER: US 09/105,839  
 ; PRIOR FILING DATE: 1998-06-26  
 ; PRIOR APPLICATION NUMBER: US 08/851,130  
 ; PRIOR FILING DATE: 1997-05-05  
 ; NUMBER OF SEQ ID NOS: 132  
 ; SEQ ID NO 2  
 ; LENGTH: 931  
 ; TYPE: DNA  
 ; ORGANISM: Homo sapiens  
 US-10-177-277-2

Query Match 86.2%; Score 496.6; DB 14; Length 931;  
 Best Local Similarity 91.5%; Pred. No. 1.1e-148;  
 Matches 526; Conservative 0; Mismatches 49; Indels 0; Gaps 0;  
 QY 1 ATGAACGAGAGCGCTTTGCAAGGAGACCCAGGAGTATGCTCAAAATATCAGAGAAG 60  
 Db 99 ATGAACGAGAGCGCTTTGCAAGGAGACCCAGGAGTATGCTCAAAATATCAGAGAAG 158

```
QY 61 TTACGAAAGCCCTTCGATGATATTCGCAAAATACCTTCTTAAGAAAGAGTGGGAAAAGATG 120
Db 159 ATCCAAAAGCCCTTCGATGATATTCGCAAAATACCTTCTTAAGAAAGAGTGGGAAAAGATG 218
QY 121 AATCTCTCGAGAAATCGTCTATGTGTATATGAAGCTAACTATGAGGTCTAGCTAAA 180
Db 219 AAAGCCCTCGAGAAATCTTCTATGTGTATATGAAGAAAGATGAGGCTATGACTAAA 278
QY 181 CTAGGTTTCAAGGTCAACCTCCACCTTTCATCGTGTAGTAAACGGGCTGCAGACTTCCAC 240
Db 279 CTAGGTTTCAAGGCCACCTCCACCTTTCATGTGTATATGAAGCTAACTATGAGGTCTAGCTAAA 338
QY 241 GCGAATGATTTGGTAAGTGGCAGAGGATCTGCGCCCAAAATGATGGGAAAAGAGTGG 300
Db 339 GCGAATGATTTGGTAAGTGGCAGAGGATCTGCGCCCAAAATGATGGGAAAAGAGTGG 398
QY 301 TTCCGAGGCTCCAGAGATCTTCCGGAAGATCATGCCCAAGAGTCCCAAGAGCCAGAGGAGAA 360
Db 459 AATGATTCGGAGGAGTGGCAGAGGATCTGCGCCCAAAATGATGGGAAAAGAGTGG 518
QY 421 CCCCCGGGAAATCCAGTACCTTGGAGAGATTAACAAGACATCTGGACCCAAAAGGGGG 480
Db 519 CCCCCGGGAAATCCAGTACCTTGGAGAGATTAACAAGACATCTGGACCCAAAAGGGGG 578
QY 481 AATGATTCGGAGGAGTGGCAGAGGATCTGCGCCCAAAATGATGGGAAAAGAGTGG 540
Db 579 GAACATGCTGGACCCACAGACTGCGTGGAGAGAAACAGCTGTGTATGAAGAGATC 638

RESULT 10
US-09-954-531-1000
; Sequence 1000, Application US/09954531
; Patent No. US20020165180A1
; GENERAL INFORMATION:
; APPLICANT: Weaver, Zoe
; TITLE OF INVENTION: Process for Identifying Anti-Cancer Therapeutic Agents Using Cand
; FILE OF INVENTION: Gene Sets
; FILE REFERENCE: 689290-77
; CURRENT APPLICATION NUMBER: US/09/954,531
; PRIOR FILING DATE: 2002-05-02
; PRIOR APPLICATION NUMBER: US/60/233,133
; PRIOR FILING DATE: 2000-09-18
; PRIOR APPLICATION NUMBER: US/60/234,009
; PRIOR FILING DATE: 2000-09-20
; PRIOR APPLICATION NUMBER: US/60/234,034
; PRIOR FILING DATE: 2000-09-22
; PRIOR APPLICATION NUMBER: US/60/234,567
; NUMBER OF SEQ ID NOS: 1392
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 1000
; LENGTH: 1309
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)...(1309)
; OTHER INFORMATION: n=a,t,g or c
US-09-954-531-1000

Query Match 86.2%; Score 496.6; DB 9; Length 1309;
Best Local Similarity 91.5%; Pred. No. 1.4e-148;
Matches 526; Conservative 0; Mismatches 49; Indels 0; Gaps 0;
```

```
QY 1 ATGAACGGAGAGCAGCCCTTTCGAGGAGACCCAGGAGTGCCTCAATATCAGAGAAG 60
Db 99 ATGAACGGAGAGCAGCCCTTTCGAGGAGACCCAGGAGTGCCTCAATATCAGAGAAG 158
QY 61 TTACGAAAGCCCTTCGATGATATTCGCAAAATACCTTCTTAAGAAAGAGTGGGAAAAGATG 120
Db 159 ATCCAAAAGCCCTTCGATGATATTCGCAAAATACCTTCTTAAGAAAGAGTGGGAAAAGATG 218
QY 121 AATCTCTCGAGAAATCGTCTATGTGTATATGAAGCTAACTATGAGGTCTAGCTAAA 180
Db 219 AAAGCCCTCGAGAAATCTTCTATGTGTATATGAAGAAAGATGAGGCTATGACTAAA 278
QY 181 CTAGGTTTCAAGGTCAACCTCCACCTTTCATCGTGTAGTAAACGGGCTGCAGACTTCCAC 240
Db 279 CTAGGTTTCAAGGCCACCTCCACCTTTCATGTGTATATGAAGCTAACTATGAGGTCTAGCTAAA 338
QY 241 GCGAATGATTTGGTAAGTGGCAGAGGATCTGCGCCCAAAATGATGGGAAAAGAGTGG 300
Db 339 GCGAATGATTTGGTAAGTGGCAGAGGATCTGCGCCCAAAATGATGGGAAAAGAGTGG 398
QY 301 TTCCGAGGCTCCAGAGATCTTCCGGAAGATCATGCCCAAGAGTCCCAAGAGCCAGAGGAGAA 360
Db 399 TTCCGAGGCTCCAGAGATCTTCCGGAAGATCATGCCCAAGAGTCCCAAGAGCCAGAGGAGAA 458
QY 361 AATGATTCGGAGGAGTGGCAGAGGATCTGCGCCCAAAATGATGGGAAAAGAGTGG 420
Db 459 AATGATTCGGAGGAGTGGCAGAGGATCTGCGCCCAAAATGATGGGAAAAGAGTGG 518
QY 421 CCCCCGGGAAATCCAGTACCTTGGAGAGATTAACAAGACATCTGGACCCAAAAGGGGG 480
Db 519 CCCCCGGGAAATCCAGTACCTTGGAGAGATTAACAAGACATCTGGACCCAAAAGGGGG 578
QY 481 AATGATTCGGAGGAGTGGCAGAGGATCTGCGCCCAAAATGATGGGAAAAGAGTGG 540
Db 579 GAACATGCTGGACCCACAGACTGCGTGGAGAGAAACAGCTGTGTATGAAGAGATC 638
QY 541 AGCGACCCCTGAGGAGAGTGCAGAGTAACTCCCTC 575
Db 639 AGCGACCCCTGAGGAGAGTGCAGAGTAACTCCCTC 673

RESULT 11
US-09-954-531-1392
; Sequence 1392, Application US/09954531
; Patent No. US20020165180A1
; GENERAL INFORMATION:
; APPLICANT: Weaver, Zoe
; TITLE OF INVENTION: Process for Identifying Anti-Cancer Therapeutic Agents Using Car
; FILE OF INVENTION: Gene Sets
; FILE REFERENCE: 689290-77
; CURRENT APPLICATION NUMBER: US/09/954,531
; PRIOR FILING DATE: 2002-05-02
; PRIOR APPLICATION NUMBER: US/60/233,133
; PRIOR FILING DATE: 2000-09-18
; PRIOR APPLICATION NUMBER: US/60/234,009
; PRIOR FILING DATE: 2000-09-20
; PRIOR APPLICATION NUMBER: US/60/234,034
; PRIOR FILING DATE: 2000-09-22
; PRIOR APPLICATION NUMBER: US/60/234,509
; PRIOR FILING DATE: 2000-09-22
; PRIOR APPLICATION NUMBER: US/60/234,567
; NUMBER OF SEQ ID NOS: 1392
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 1392
; LENGTH: 1309
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)...(1309)
; OTHER INFORMATION: n=a,t,g or c
```

US-09-954-531-1392

```

Query Match      86.2%; Score 496.6; DB 9; Length 1309;
Best Local Similarity 91.5%; Pred. No. 1.4e-148;
Matches 526; Conservative 0; Mismatches 49; Indels 0; Gaps 0;

QY 1 ATGAACGGAGAGCAGCCCTTCCAGAGGAGACCCAGGATGATGCTCAATATACAGAGAG 60
DB 99 ATGAACGGAGAGCAGCCCTTCCAGAGGAGACCCAGGATGATGCTCAATATACAGAGAG 158
QY 61 TTACGAAAGGCGCTTCGATGATATGCAAAATCTCTTAAGAAAGAGTGGGAAAGATG 120
DB 159 ATCCAAAGGCGCTTCGATGATATGCAAAATCTCTTAAGAAAGAGTGGGAAAGATG 218
QY 121 AAATCCTCGGAGAAATCGTCTATGTATATGAAGCTAAACTATGAGGTCTATGACTAAA 180
DB 219 AAAGCTCGGAGAAATCTCTATGTATATGAAGCTAAACTATGAGGTCTATGACTAAA 278
QY 181 CTAGGTTTCAAGGTCAACCTCCACCTTTCATGCGTAGTAAACGGGCTCGAGACTTCCAC 240
DB 279 CTAGGTTTCAAGGCGCCCTTCCACCTTTCATGTATATGAAGCTAAACTATGAGGTCTATGACTAAA 338
QY 241 GGAATGATTTTGGTAAAGATGCAAAACCCAGAGATCAGCTTGAACGTCTCTCAGATGACT 300
DB 339 GGAATGATTTTGGTAAAGATGCAAAACCCAGAGATCAGCTTGAACGTCTCTCAGATGACT 398
QY 301 TTCCGAGGCTCCAGAGATCTTCCGAGAGATCATGCCCAAGAGCCAGCAGAGAGAGAA 360
DB 399 TTCCGAGGCTCCAGAGATCTTCCGAGAGATCATGCCCAAGAGCCAGCAGAGAGAGAA 458
QY 361 AATGTTTGAAGAGAGTGCAGAGGATCTGCCCAACAAATGATGGGAAACAGCTGTGC 420
DB 459 AATGTTTGAAGAGAGTGCAGAGGATCTGCCCAACAAATGATGGGAAACAGCTGTGC 518
QY 421 CCCCCGGGAAATCCAAAGTACCTTGGAGAGATTAACAAGACATCTGCAACCCAAAGGGGG 480
DB 519 CCCCCGGGAAATCCAAAGTACCTTGGAGAGATTAACAAGACATCTGCAACCCAAAGGGGG 578
QY 481 AAATGCTTGGAGCCACAGATCTGCTGAGAGAGAGAGAGAGAGAGATC 540
DB 579 GAACATGCTTGGAGCCACAGATCTGCTGAGAGAGAGAGAGAGAGATC 638
QY 541 AGCAGCCCTGAGAGAGATGACGAGTACTCCCTC 575
DB 639 AGCAGCCCTGAGAGAGATGACGAGTACTCCCTC 673

RESULT 12
US-10-159-563-402
; Sequence 402, Application US/10159563
; Publication No. US20040009154A1
; GENERAL INFORMATION:
; APPLICANT: Khan, Javed
; APPLICANT: Ringner, Markus
; APPLICANT: Peterson, Carsten
; APPLICANT: Meltzer, Paul
; TITLE OF INVENTION: SELECTIONS OF GENES AND METHODS OF USING THE SAME FOR
; TITLE OF INVENTION: DIAGNOSIS AND FOR TARGETING THE THERAPY OF SELECT CANCERS
; FILE REFERENCE: 11613.56US11
; CURRENT FILING DATE: 2002-12-09
; PRIOR FILING DATE: 2002-04-25
; NUMBER OF SEQ ID NOS: 444
; SOFTWARE: Patent in version 3.1
; LENGTH: 1249
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-159-563-402
Query Match      86.1%; Score 496; DB 15; Length 1249;
Best Local Similarity 91.3%; Pred. No. 2.1e-148;

```

```

Matches 526; Conservative 0; Mismatches 50; Indels 0; Gaps 0;

QY 1 ATGAACGGAGAGCAGCCCTTCCAGAGGAGACCCAGGATGATGCTCAATATACAGAGAG 60
DB 54 ATGAACGGAGAGCAGCCCTTCCAGAGGAGACCCAGGATGATGCTCAATATACAGAGAG 113
QY 61 TTACGAAAGGCGCTTCGATGATATGCAAAATCTCTTAAGAAAGAGTGGGAAAGATG 120
DB 114 ATCAAAAGGCGCTTCGATGATATGCAAAATCTCTTAAGAAAGAGTGGGAAAGATG 173
QY 121 AAATCCTCGGAGAAATCGTCTATGTATATGAAGCTAAACTATGAGGTCTATGACTAAA 180
DB 174 AAAGTCTCGGAGAAATCGTCTATGTATATGAAGCTAAACTATGAGGTCTATGACTAAA 233
QY 181 CTAGGTTTCAAGGTCAACCTCCACCTTTCATGCGTAGTAAACGGGCTCGAGACTTCCAC 240
DB 234 CTAGGTTTCAAGGCGCCCTTCCACCTTTCATGCGTAGTAAACGGGCTCGAGACTTCCAC 293
QY 241 GGAATGATTTTGGTAAACCGATCGAAACCCAGAGATCAGCTTGAACGTCTCTCAGATGACT 300
DB 294 GGAATGATTTTGGTAAACCGATCGAAACCCAGAGATCAGCTTGAACGTCTCTCAGATGACT 353
QY 301 TTCCGAGGCTCCAGAGATCTTCCGAGAGATCATGCCCAAGAGCCAGCAGAGAGAGAA 360
DB 354 TTCCGAGGCTCCAGAGATCTTCCGAGAGATCATGCCCAAGAGCCAGCAGAGAGAGAA 413
QY 361 AATGTTTGAAGAGAGTGCAGAGGATCTGCCCAACAAATGATGGGAAACAGCTGTGC 420
DB 414 AATGTTTGAAGAGAGTGCAGAGGATCTGCCCAACAAATGATGGGAAACAGCTGTGC 473
QY 421 CCCCCGGGAAATCCAAAGTACCTTGGAGAGATTAACAAGACATCTGCAACCCAAAGGGGG 480
DB 474 CCCCCGGGAAATCCAAAGTACCTTGGAGAGATTAACAAGACATCTGCAACCCAAAGGGGG 533
QY 481 AAATGCTTGGAGCCACAGATCTGCTGAGAGAGAGAGAGAGAGATC 540
DB 534 GAACATGCTTGGAGCCACAGATCTGCTGAGAGAGAGAGAGAGATC 593
QY 541 AGCAGCCCTGAGAGAGATGACGAGTAACTCCCTC 576
DB 594 AGCATCTCTGAGAGAGATGATGATTAACCTCCCTC 629

RESULT 13
US-09-975-856-2
; Sequence 2, Application US/09975856
; Publication No. US20030023057A1
; GENERAL INFORMATION:
; APPLICANT: Gure, Ali O.; Tureci, Ozlem; Sahin, Ugur; Tsang, Solan;
; Scanlan, Matthew J.; Knuth Alexander; Pfreundschuh, Michael; Old,
; Lloyd J.; Chen, Yao-Tseng
; TITLE OF INVENTION: Isolated Nucleic Acid Molecules Encoding SSX
; Family
; Members And Uses Thereof
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fulbright & Jaworski L.L.P.
; STREET: 666 Fifth Avenue
; CITY: New York City
; STATE: New York
; ZIP: 10103
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.5 inch, 144 kb storage
; COMPUTER: IBM
; OPERATING SYSTEM: PC-DOS
; SOFTWARE: Wordperfect
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/975,856
; FILING DATE: 11-Oct-2001
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 09/796,780
; FILING DATE: 2001-03-01

```

ATTORNEY/AGENT INFORMATION:  
NAME: Hanson, No. US20030023057Alman D.  
REGISTRATION NUMBER: 30,946  
REFERENCE/DOCKET NUMBER: LUD 5480  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (212) 318-3168  
TELEFAX: (212) 752-5958  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 576 nucleotides  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
SEQUENCE DESCRIPTION: SEQ ID NO: 2:  
US-09-975-856-2

Query Match 83.7%; Score 482.2; DB 10; Length 576;  
Best Local Similarity 89.9%; Pred. No. 3.7e-144;  
Matches 517; Conservative 0; Mismatches 58; Indels 0; Gaps 0;

QY 1 ATGAACGGAGACGACGCTTTGCAAGGAGACCCAGGATGATGCTCAAAATATCAGAGAAG 60  
Db 1 ATGAACGGAGACGACGCTTTGCAAGGAGACCCAGGATGATGCTCAAAATATCAGAGAAG 60

QY 61 TTACGAAGGCTTCGATGATATGCGCAATATCTCTTAAGAAAGTGGGAAAGATG 120  
Db 61 ATGCAAAAGGCTTCGATGATATGCGCAATATCTCTTAAGAAAGTGGGAAAGATG 120

QY 121 AAATCCTCGGAGAAATCGTGTATGTATGAAGCTAAATATGAGGTCATGACTAAA 180  
Db 121 AAAGCTTCGAGAAATCATCTATGTATATGAAGAAAGTATGAGGCCATGACTAAA 180

QY 181 CTAGGTTTCAAGTCCACCTTCCAGGATCATGCGTAGTAAACGGCTGCAGACTTCCAC 240  
Db 181 CTAGGTTTCAAGGACCACTTCCAGGATCATGCGTAGTAAACGGCTGCAGACTTCCAG 240

QY 241 GGAATGATTTGGTAAACGATCGAACCACAGGAATCAGGTTGAACGTCCTCAGATGACT 300  
Db 241 GGAATGATTTGAATGAATACCTTAACCTGGAATCAGGTTGAACATCTCAGATGACT 300

QY 301 TTCCGAGCCTCCAGAGAAATCTTCCGAGAGATCATGCCAAGAGCCAGCAGAGGAAGAA 360  
Db 301 TTCCGAGGCTCCAGGAAATCTTCCGAGAGATCATGCCAAGAGCCAGCAGAGGAAGAA 360

QY 361 AATGTTTGAAGAAAGTGCAGAGGATCTGGCCCAAAATATGATGGGAAACAGCTGTGC 420  
Db 361 AATGATTCAAAGGGAGTGCAGAGGATCTGGCCCAAAATATGATGGGAAACAGCTGTGC 420

QY 421 CCCCCGGGAAATCCAAAGTACCTTGGAGAGATTAACAAGACATCTGGACCCAAAGGGGG 480  
Db 421 CCTCAGGAAACTTAATACCTCTGAGAGGTTTAAAGACATCTGGACCCAAAGGGGG 480

QY 481 AAACATGCTTGGACCCACAGACTGCGTGAGAGAAAGCAGCTGGTGGTTTATGAAGATC 540  
Db 481 AAACATGCTTGGACCCACAGACTGCGTGAGAGAAAGCAGCTGGTGGTTTATGAAGATC 540

QY 541 AGCGACCTCAGGAAGATGACGAGTAACCTCCCTC 575  
Db 541 AGCGACCTCAGGAAGATGACGAGTAACCTCCCTC 575

RESULT 14  
US-09-833-039-6  
; Sequence 6, Application US/09833039  
; Publication No. US20030175960A1  
; GENERAL INFORMATION:  
; APPLICANT: Tureci, Ozlem  
; APPLICANT: Sahin, Ugur  
; APPLICANT: Pfreundschuh, Michael  
; TITLE OF INVENTION: Tumor Associated Peptide and Uses Thereof  
; FILE REFERENCE: LUD 5622.1  
; CURRENT APPLICATION NUMBER: US/09/833,039  
; CURRENT FILING DATE: 2001-04-12

; PRIOR APPLICATION NUMBER: US 09/409,455  
; PRIOR FILING DATE: 1999-09-30  
; PRIOR APPLICATION NUMBER: US 09/344,040  
; PRIOR FILING DATE: 1999-06-25  
; PRIOR APPLICATION NUMBER: US 09/105,839  
; PRIOR FILING DATE: 1998-06-26  
; PRIOR APPLICATION NUMBER: US 08/851,130  
; PRIOR FILING DATE: 1997-05-05  
; NUMBER OF SEQ ID NOS: 129  
; SEQ ID NO 6  
; LENGTH: 576  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-09-833-039-6

Query Match 83.7%; Score 482.2; DB 10; Length 576;  
Best Local Similarity 89.9%; Pred. No. 3.7e-144;  
Matches 517; Conservative 0; Mismatches 58; Indels 0; Gaps 0;

QY 1 ATGAACGGAGACGACGCTTTGCAAGGAGACCCAGGATGATGCTCAAAATATCAGAGAAG 60  
Db 1 ATGAACGGAGACGACGCTTTGCAAGGAGACCCAGGATGATGCTCAAAATATCAGAGAAG 60

QY 61 TTACGAAGGCTTCGATGATATGCGCAATATCTCTTAAGAAAGTGGGAAAGATG 120  
Db 61 ATGCAAAAGGCTTCGATGATATGCGCAATATCTCTTAAGAAAGTGGGAAAGATG 120

QY 121 AAATCCTCGGAGAAATCGTGTATGTATGAAGCTAAATATGAGGTCATGACTAAA 180  
Db 121 AAAGCTTCGAGAAATCATCTATGTATATGAAGAAAGTATGAGGCCATGACTAAA 180

QY 181 CTAGGTTTCAAGTCCACCTTCCAGGATCATGCGTAGTAAACGGCTGCAGACTTCCAC 240  
Db 181 CTAGGTTTCAAGGACCACTTCCAGGATCATGCGTAGTAAACGGCTGCAGACTTCCAG 240

QY 241 GGAATGATTTGGTAAACGATCGAACCACAGGAATCAGGTTGAACGTCCTCAGATGACT 300  
Db 241 GGAATGATTTGAATGAATACCTTAACCTGGAATCAGGTTGAACATCTCAGATGACT 300

QY 301 TTCCGAGCCTCCAGAGAAATCTTCCGAGAGATCATGCCAAGAGCCAGCAGAGGAAGAA 360  
Db 301 TTCCGAGGCTCCAGGAAATCTTCCGAGAGATCATGCCAAGAGCCAGCAGAGGAAGAA 360

QY 361 AATGTTTGAAGAAAGTGCAGAGGATCTGGCCCAAAATATGATGGGAAACAGCTGTGC 420  
Db 361 AATGATTCAAAGGGAGTGCAGAGGATCTGGCCCAAAATATGATGGGAAACAGCTGTGC 420

QY 421 CCCCCGGGAAATCCAAAGTACCTTGGAGAGATTAACAAGACATCTGGACCCAAAGGGGG 480  
Db 421 CCTCAGGAAACTTAATACCTCTGAGAGGTTTAAAGACATCTGGACCCAAAGGGGG 480

QY 481 AAACATGCTTGGACCCACAGACTGCGTGAGAGAAAGCAGCTGGTGGTTTATGAAGATC 540  
Db 481 AAACATGCTTGGACCCACAGACTGCGTGAGAGAAAGCAGCTGGTGGTTTATGAAGATC 540

QY 541 AGCGACCTCAGGAAGATGACGAGTAACCTCCCTC 575  
Db 541 AGCGACCTCAGGAAGATGACGAGTAACCTCCCTC 575

RESULT 15  
US-10-177-277-6  
; Sequence 6, Application US/10177277  
; Publication No. US20030185844A1  
; GENERAL INFORMATION:  
; APPLICANT: Tureci, Ozlem  
; APPLICANT: Sahin, Ugur  
; APPLICANT: Pfreundschuh, Michael  
; APPLICANT: Ramensee, Hans Georg  
; APPLICANT: Stewanovic, Stefan  
; TITLE OF INVENTION: Method for Determining Presence of Cancer in a Sample By Determ:  
; FILE REFERENCE: LUD 5622.1  
; CURRENT APPLICATION NUMBER: Expression of an Ssx Gene, Peptides Derived From Said Ssx Gene  
; CURRENT FILING DATE: Gene, and Uses Thereof

```
; FILE REFERENCE: LUD 5556.1
; CURRENT APPLICATION NUMBER: US/10/177,277
; CURRENT FILING DATE: 2002-06-21
; PRIOR APPLICATION NUMBER: US/09/344,040
; PRIOR FILING DATE: 1999-06-25
; PRIOR APPLICATION NUMBER: US 09/105,839
; PRIOR FILING DATE: 1998-06-26
; PRIOR APPLICATION NUMBER: US 08/851,130
; PRIOR FILING DATE: 1997-05-05
; NUMBER OF SEQ ID NOS: 132
; SEQ ID NO 6
; LENGTH: 576
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-177-277-6

Query Match      83.7%; Score 482.2; DB 14; Length 576;
Best Local Similarity 89.9%; Pred. No. 3.7e-144;
Matches 517; Conservative 0; Mismatches 58; Indels 0; Gaps 0;

QY      1 ATGAACGGAGACGACGCCCTTTCAGAGGAGACCCAGGGATGATGCTCAAATATCCAGAGAAG 60
Db      |||
QY      1 ATGAACGGAGAGCAGACGCCCTTGTACGGAGACCTAGGGTGGTTCTCAATATCCACAGAAG 60
Db      |||
QY      61 TTACGAAAGGCGCTTCGATGATATGCAAAATCTTCTTAAGAAAGAGTGGGAAAAGATG 120
Db      |||
QY      61 ATGCAAAAGGCGCTTCGATGATATGCAAAATCTTCTCTGAGAAAGAGTGGGAAAAGATG 120
Db      |||
QY      121 AAATCTCCGAGAAATCTCTATGTATATGAAGCTAAACTATCAGGTCTGACTAAA 180
Db      |||
QY      121 AAAGCCTCGGAGAAATCTATGTATATGAAGAGAGATGAGGCCATGACTAAA 180
Db      |||
QY      181 CTAGGTTTCAAGTCAACCTCCACACCTTTCATGCGTAGTAAACGGGCTGCAGACTTCCAC 240
Db      |||
QY      181 CTAGGTTTCAAGGCAACCTCCACACCTTTCATGCGTAGTAAACGGGCTGCAGACTTCCAG 240
Db      |||
QY      241 GGGATGATTTGGTAAACGATCGAAACCAAGGAATCAGGTTGAACGTCCTCAGATGACT 300
Db      |||
QY      241 GGGATGATTTTGAATGATGACCTAACCGTGGGAATCAGGTTGAACATCCTCAGATGACT 300
Db      |||
QY      301 TTCGGCAGGCTCCAGAGATCTTCCGAGAGATCATGCCCAAGAGCCCAAGAGCCAGAGAGAGAA 360
Db      |||
QY      301 TTCGGCAGGCTCCAGGATCTTCCGAGAGATCATGCCCAAGAGCCCAAGAGCCAGAGAGAGAA 360
Db      |||
QY      361 AATGTTTGAAGGAAGTGCAGAGGATCTGGCCCAACAAATGATGGGAAACAGCTGTGC 420
Db      |||
QY      361 AATGTTTGAAGGAAGTGCAGAGGATCTGGCCCAACAAATGATGGGAAACAGCTGTGC 420
Db      |||
QY      421 CCCCAGGGAATCCAGTACCTTGGAGAGATTAAACAGACATCTTGACCCCAAGGGGG 480
Db      |||
QY      421 CCTCAGGAAACTAAATACCTCTGAGAAAGTTAACAGACATCTGGACCCCAAGGGGG 480
Db      |||
QY      481 AAACATGCTGGACCCACAGATCTGCGTAGAGAAAGCAGCTGGTGTATGAAGATC 540
Db      |||
QY      481 AAACATGCTGGACCCACAGATCTGCGTAGAGAAAGCAGCTGGTGTATGAAGATC 540
Db      |||
QY      541 AGCAGCCCTGAGGAAGATGACGAGTAACCTCCCTC 575
Db      |||
QY      541 AGCAGCCCTGAGGAAGATGACGAGTAACCTCCCTC 575
Db      |||
```

Search completed: March 31, 2004, 18:32:55  
Job time : 402 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - protein search, using frame\_plus\_n2p model

Run on: March 31, 2004, 13:39:56 ; Search time 14 Seconds  
(without alignments)  
4284.630 Million cell updates/sec

Title: US-09-975-856-1

Perfect score: 1025

Sequence: 1 ATGACGAGGAGACGACCTT.....ATGACGAGTAACCTCCCTCG 576

Scoring table: BLOSUM62

Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 283362

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-Q=/cgn2\_1/USPTO\_spool\_p/US09975856/runat\_31032004\_132903\_14738/app\_query.fasta\_1.775  
-DB=SwissProt\_42 -Qfmt=fastan -SUFFIX=xsp -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0  
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LISF=45  
-DOCALIGN=200 -THR\_SCORE=pct -THR\_MAX=100 -THR\_MIN=0 -ALIGN=15 -MODE=LOCAL  
-OUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US09975856@cgn\_1\_1\_16 @runat\_31032004\_132903\_14738 -NCPU=3  
-NO MMAP -LARGEQUERY -NEG SCORE=0 -WAIT -DSPBLOCK=100 -LONGLOG  
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THRADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : SwissProt\_42.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	1004	98.0	188	1 SSX4_HUMAN	O60224 homo sapien
2	801	78.1	188	1 SSX3_HUMAN	Q99909 homo sapien
3	798	77.9	188	1 SSX2_HUMAN	Q16385 homo sapien
4	784	76.5	188	1 SSX1_HUMAN	Q16384 homo sapien
5	780	76.1	188	1 SSX5_HUMAN	O60225 homo sapien
6	113.5	11.1	289	1 ZN75_HUMAN	P51815 homo sapien
7	101	9.9	642	1 Z398_HUMAN	Q8cd17 homo sapien
8	94	9.2	595	1 Z317_HUMAN	O96p35 homo sapien
9	89.5	8.9	1505	1 PK37_RAT	O70173 rattus norv
10	88.5	8.8	531	1 FM01_CANFA	O541a2 canis famul
11	88	8.6	743	1 REP1_MOUSE	Q54916 mus musculu
12	87.5	8.7	531	1 FM01_PIG	P16549 sus scrofa
13	87	8.5	744	1 REP1_HUMAN	Q46d71 homo sapien
14	86.5	8.4	396	1 SX11_CHICK	P98435 gallus galli
15	86.5	8.4	446	1 ZN38_HUMAN	P17036 homo sapien
16	86.5	8.4	488	1 ZF92_MOUSE	Q62196 mus musculu
17	86.5	8.4	572	1 TC17_MOUSE	Q61751 mus musculu
18	86	8.4	480	1 WR61_ARATH	Q8wvv6 arabidopsis

19	86	8.4	834	1 CASL_HUMAN	Q14511 homo sapien
20	85.5	8.3	680	1 Z334_HUMAN	Q9hcz1 homo sapien
21	85.5	8.3	738	1 ZN84_HUMAN	P51323 homo sapien
22	84.5	8.2	429	1 H18X_METJA	Q5851 methanococc
23	84.5	8.2	463	1 Z331_HUMAN	Q9nqx6 homo sapien
24	84.5	8.4	534	1 FM01_RABIT	P17636 oryctolagus
25	84	8.2	524	1 Z479_HUMAN	Q96jca4 homo sapien
26	83.5	8.1	636	1 ZF90_MOUSE	Q61567 mus musculu
27	83.5	8.1	1102	1 ADNP_HUMAN	Q9h2p0 homo sapien
28	83.5	8.1	1630	1 MSP1_PLAEX	P04332 plasmodium
29	83.5	8.1	1639	1 MSP1_PLAFW	P04333 plasmodium
30	82.5	8.0	553	1 Z334_HUMAN	O75467 homo sapien
31	82.5	8.0	643	1 ZN74_HUMAN	Q16587 homo sapien
32	82.5	8.0	1035	1 RRPO_BMYVF	P09507 beet wester
33	81.5	8.0	1035	1 Z205_HUMAN	O95201 homo sapien
34	81.5	8.0	504	1 GLOX_SOYBN	P11827 glycine max
35	81.5	8.0	639	1 SP41_YEAST	P38904 saccharomyc
36	81.5	8.0	1395	1 DPO3_MYCPN	P75080 mycoplasma
37	81	7.9	458	1 ZN19_HUMAN	P17023 homo sapien
38	81	7.9	751	1 Z337_HUMAN	Q9y3m9 homo sapien
39	81	8.0	933	1 PERT_CANFA	Q8hyb7 canis famul
40	80.5	7.9	439	1 ZN10_HUMAN	P21506 homo sapien
41	80.5	7.9	470	1 Z436_HUMAN	Q9c0f3 homo sapien
42	80.5	7.9	659	1 Z304_HUMAN	Q9hcz3 homo sapien
43	80.5	7.9	998	1 GTFI_HUMAN	P78347 h general t
44	79.5	7.9	531	1 FM01_HUMAN	Q01740 homo sapien
45	79.5	7.8	626	1 Z471_HUMAN	Q9bx82 homo sapien

## ALIGNMENTS

RESULT 1  
SSX4\_HUMAN  
ID SSX4\_HUMAN STANDARD; PRT; 188 AA.  
AC O60224; Q9UJY9;  
DT 15-JUL-1999 (Rel. 38, Created)  
DT 15-JUL-1999 (Rel. 38, Last sequence update)  
DT 10-OCT-2003 (Rel. 42, Last annotation update)  
DE SSX4 protein.  
GN SSX4.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Euthera; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=98021352; PubMed=9378559;  
RA Gure A.O., Tuercio O., Sahin U., Tsang S., Scanlan M.J., Jager E.,  
RA Knuth A., Pfeundschn M., Old L.J., Chen Y.-T.;  
RT "SSX: a multigene family with several members transcribed in normal  
RT testis and human cancer."  
RL Int. J. Cancer 72:965-971(1997).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Urinary bladder;  
RX MEDLINE=22388257; PubMed=12477932;  
RA Strauberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,  
RA Klausner R.D., Collins F.S., Wagner L., Sherman C.M., Schuler G.D.,  
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,  
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,  
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,  
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,  
RA Brownstein M.J., Udell T.B., Toshiyuki S., Carninci P., Prange C.,  
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,  
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,  
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,  
RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,  
FAhey J., Helton E., Kettman M., Madao A., Rodrigues S., Sanchez A.,  
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,  
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,  
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,  
RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smallus D.E.,  
RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;

RT "Generation and initial analysis of more than 15,000 full-length  
 RL human and mouse cDNA sequences.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).  
 RN [3]  
 RP SEQUENCE OF 1-155 FROM N.A.  
 RA Bleischmidt K., Nyakatura G., Strom T.M., Drescher B., Menzel U.,  
 RA Meindl A., Rosenthal A.;  
 RL Submitted (OCT-1999) to the EMBL/GenBank/DBJ databases.  
 CC -!- FUNCTION: Could act as a modulator of transcription.  
 CC -!- SIMILARITY: Belongs to the SSX family.  
 CC -!- SIMILARITY: Contains 1 KRAB-related domain.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL; U90841; AAC05820.1; -;  
 DR EMBL; BC005325; AAH05325.1; -;  
 DR EMBL; AF196972; AAF06796.1; -;  
 DR Genew; HGNC:11338; SSX4.  
 DR MIM; 300326;  
 DR InterPro; IPR001909; KRAB.  
 DR InterPro; IPR003655; KRAB\_related.  
 DR Pfam; PF01352; KRAB; 1.  
 DR SMART; SM00349; KRAB; 1.  
 DR PROSITE; PS50806; KRAB RELATED; 1.  
 KW Multigene family; Transcription regulation.  
 FT DOMAIN 20 83  
 FT KRAB-RELATED.  
 SQ SEQUENCE 188 AA; 21858 MW; 4ACA2A8737507AE5 CRC64;

# Alignment Scores:

Pred. No.: 1-37e-91 Length: 188  
 Score: 1004.00 Matches: 188  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 97.95% Indels: 0  
 DB: 1 Gaps: 0

US-09-975-856-1 (1-576) x SSX4\_HUMAN (1-188)

QY 1 ATGAACGAGACGACGCTTGGACGAGACCCAGGATGATCTCAATATCAGAGAG 60  
 DB 1 MetAsnGlyAspAlaPheAlaArgProArgAspAlaGlnileSerGluLys 20  
 QY 61 TTACGAAAGGCTTCGATGATATTCGCAATCTCTCTAAGAAAGAGTGGGAAAGATG 120  
 DB 21 LeuArgLysAlaPheAspAlaLysTyrPheSerLysLysGluTrpGluLysMet 40  
 QY 121 AAATCTCCGAGAAATCGTCTATGTATGATATGAAGCTAACTATCAGGTCTAGCTAA 180  
 DB 41 LysSerSerGluLysileValTyrValTyrMetLysLeuAsnTyrGluValMetThrLys 60  
 QY 181 CTAGTTTCAAGTTCACCTCCACCTTCATCGTAGTAAGCGGCTGCAGACTTCCAC 240  
 DB 61 LeuGlyPheLysValThrLeuProPheMetArgSerLysArgAlaAlaaspPheHis 80  
 QY 241 GGAAGTATTGGTGAACCATGACCAACACAGCAATCAGTTGAAGTCTCAGTACT 300  
 DB 81 GlyAsnaspPheGlyAsnaspArgAsnHisArgAsnGlnValGluArgProGlnMetThr 100  
 QY 301 TTCGGCAGCTCCAGAGATCTTCCGAGATCATGCCAAGAGCCAGCAGAGAGAA 360  
 DB 101 PheGlySerLeuGlnArgilePheProLysileMetProLysLysProAlaGluGlu 120  
 QY 361 ATGTTTGAAGAGTGCAGAGGATCTGGCCCAAAATATGATGGGAAACAGCTGTGC 420  
 DB 121 AsnGlyLeuLysGluValProGluAlaSerGlyProGlnAsnAspGlyLysGlnLeuCys 140  
 QY 421 CCCCCGGGAATCCAGTACCTTGGAGAGATTAAACAGACATCTCGACCCCAAGGGGG 480

Db 141 ProProGlyAsnProSerThrLeuGluLysileAsnLysThrSerGlyProLysArgGly 160  
 QY 481 AAACATGCTGACCCACAGACTGCTGAGACAAAGACAGCTGCTGTTTATGAGAGATC 540  
 Db 161 LysHisAlaTrpThrHisArgLeuArgGluArgLysGlnLeuValTyrGluGlu 180  
 QY 541 AGCAGCCCTGAGGAGAGATGACGAG 564  
 Db 181 SerAspProGluGluAspAspGlu 188

## RESULT 2

SSX3\_HUMAN  
 ID SSX3\_HUMAN STANDARD; PRT; 188 AA.  
 AC Q99309; O60223;  
 DT 15-DEC-1998 (Rel. 37, Created)  
 DT 15-DEC-1998 (Rel. 37, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE SSX3 protein.  
 GN SSX3.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Fibrosarcoma, and Testis;  
 RX MEDLINE=96302330; PubMed=8697803;  
 RA de Leeuw B., Balenans M., Geurts van Kessel A.;  
 RT "A novel Kruppel-associated box containing the SSX gene (SSX3) on the  
 RT human X chromosome is not implicated in t(X;18)-positive synovial  
 RT sarcomas";  
 RL Cytogenet. Cell Genet. 73:179-183(1996).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC MEDLINE=98021352; PubMed=9378559;  
 RA Gure A.O., Tuerci O., Sahin U., Tsang S., Scanlan M.J., Jager E.,  
 RA Knuth A., Pfeundschn M., Old L.J., Chen Y.-T.;  
 RT "SSX: a multigene family with several members transcribed in normal  
 RT testis and human cancer";  
 RL Int. J. Cancer 72:965-971(1997).  
 CC -!- FUNCTION: Could act as a modulator of transcription.  
 CC -!- SIMILARITY: Belongs to the SSX family.  
 CC -!- SIMILARITY: Contains 1 KRAB-related domain.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL; S82471; AAB37436.2; -;  
 DR EMBL; U90840; AAC05819.1; -;  
 DR Genew; HGNC:11337; SSX3.  
 DR MIM; 300325; -;  
 DR InterPro; IPR001909; KRAB.  
 DR InterPro; IPR003655; KRAB\_related.  
 DR Pfam; PF01352; KRAB; 1.  
 DR SMART; SM00349; KRAB; 1.  
 DR PROSITE; PS50806; KRAB RELATED; 1.  
 KW Multigene family; Transcription regulation.  
 FT DOMAIN 20 83  
 FT KRAB-RELATED.  
 FT CONFLICT 95 95 L -> Q (IN REF. 2).  
 SQ SEQUENCE 188 AA; 21682 MW; 368C7487C1450E72 CRC64;

## Alignment Scores:

Pred. No.: 1-85e-71 Length: 188  
 Score: 801.00 Matches: 152  
 Percent Similarity: 84.57% Conservative: 7  
 Best Local Similarity: 80.85% Mismatches: 29  
 Query Match: 78.11% Indels: 0



```

DB:          1      1      0      Gaps:
US-09-975-856-1 (1-576) x SSX2_HUMAN (1-188)
QY 1 ATGACGGAGAGCGCCCTTCACAGGAGACCCAGGATGATGCTCAATATCAGAGAG 60
Db 1 MetAenGlyAspThrPheAlaArgArgProThrValGlyAlaGlnIleProGluLys 20
QY 61 TTACGAAAGCCCTTCATGATATGCAATATCTCTCTAAGAAAGAGTGGGAAAGATG 120
Db 21 IleGlnLysAlaPheAspAlaLysTyrPheSerLysGluGluTrpGluLysMet 40
QY 121 AAATCCTCGGAGAAATCTCTATGATATGAAGCTAAATATGAGTCTATGACTAAA 180
Db 41 LysValSerGluLysIleValTyrMetLysArgLysTyrGluAlaMetThrLys 60
QY 181 CTAGCTTTCACAGTCCACCTCCACCTTCATGCTGCTAGTAAACGGCTGCACATCCAC 240
Db 61 LeuGlyPheLysAlaIleLeuProSerPheMetArgAsnLysArgValThrAspPheGln 80
QY 241 GGAATGATATTTGGTAACGATCGAAACACACAGGAATCAGTTGAACTGCTCAGATGACT 300
Db 81 GlyAsnAspPheAspAsnAspProAsnArgGlyAsnGlnValLeuArgProGlnMetThr 100
QY 301 TTCGCGAGCTCCAGAGAAATCTCCGAGAGATCATGCCCCAGAGCCAGCAGAGAGAA 360
Db 101 PheGlyArgLeuGlnGlyIlePheProLysIleMetProLysLysProAlaGluGlu 120
QY 361 AATGTTTGAAGGAAGTGCACAGCATCTGCCACACAAATGATGGGAACAGCTGTC 420
Db 121 AsnValSerLysGluValProGluAlaSerGlyProGlnAspGlyLysGlnLeuCys 140
QY 421 CCCCCGGGAATPCAAAGTACCTTGGAGAGATTAACAAGACATCTCGACCCCAAGAGGGGG 480
Db 141 ProProGlyLysProThrThrSerGluLysIleAsnMetIleSerGlyProLysArgGly 160
QY 481 AAACATGCTGACCCACAGCATCGTGAGAGAAAGAGCTGGTGGTTATTAAGAGATC 540
Db 161 GluHisAlaTrpThrHisArgLeuArgGluArgLysGlnLeuValIleTyrGluLeu 180
QY 541 AGCGACCTCGAGGAGATGACGAG 564
Db 181 SerAspProGluGluAspGlu 188

RESULT 3
SSX2_HUMAN
ID SSX2_HUMAN STANDARD; PRT; 188 AA.
AC Q16385; Q16404; Q961P7;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE SSX2 protein (Synovial sarcoma, X breakpoint 2) (SSX) (HOM-MEL-40).
GN SSX2.
OS Homo sapiens (Human)
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
[1]
RC SEQUENCE FROM N.A.
EX MEDLINE=98292374; PubMed=7539744;
RA Crew A.J., Clark J., Fisher C., Gill S., Grimer R., Chand A.,
RA Shipley J., Gusterson B.A., Cooper C.S.;
RT "Fusion of SYT to two genes, SSX1 and SSX2, encoding proteins with
RT homology to the Kruppel-associated box in human synovial sarcoma.";
RL EMBO J. 14:2333-2340(1995).
[2]
RC SEQUENCE FROM N.A.
TX TISSUE=Skin;
RC MEDLINE=22388257; PubMed=12477932;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shemen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,

```

```

RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Lequellano N.A., Peters G.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Bosak S.A., McQuellan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Wozley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahy J., Helton E., Kettelman M., Madan A., Rodriguez S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smalhus D.E.,
RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length
RT human and mouse cDNA sequences.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).
[3]
RN SEQUENCE OF 111-188 FROM N.A.
RC TISSUE=Synovial sarcoma;
RX MEDLINE=95384157; PubMed=7655467;
RA de Leeuw B., Bailemans M., Olde Weghuis D., Geurts van Kessel A.;
RT "Identification of two alternative fusion genes, SYT-SSX1 and
RT SYT-SSX2, in t(X;18)(p11.2;q11.2)-positive synovial sarcomas.";
RL Hum. Mol. Genet. 4:1097-1099(1995).
[4]
RN SEQUENCE OF 111-188 FROM N.A. (SSX1-SSX2 FUSION PROTEIN).
RC TISSUE=Synovial sarcoma;
RX MEDLINE=95038836; PubMed=7951320;
RA Clark J., Roques P.J., Crew A.J., Gill S., Shipley J., Chan A.M.-L.,
RA Gusterson B.A., Cooper C.S.;
RT "Identification of novel genes, SYT and SSX, involved in the
RT t(X;18)(p11.2;q11.2) translocation found in human synovial sarcoma.";
RL Nat. Genet. 7:502-508(1994).
[5]
RN SEQUENCE OF 68-116 FROM N.A. (SSX1-SSX2 FUSION PROTEIN).
RX MEDLINE=96094743; PubMed=7495284;
RA Fligman I., Lonardo F., Jhanwar S.C., Gerald W.L., Woodruff J.,
RA Ladanyi M.;
RT "Molecular diagnosis of synovial sarcoma and characterization of a
RT variant SYT-SSX2 fusion transcript.";
RL Am. J. Pathol. 147:1592-1599(1995).
CC -!- FUNCTION: Could act as a modulator of transcription.
CC -!- TISSUE SPECIFICITY: Expressed at high level in the testis.
CC Expressed at low level in thyroid. Not detected in tonsil, colon,
CC lung, spleen, prostate, kidney, striated and smooth muscles.
CC Detected in rhabdomyosarcoma and fibrosarcoma cell lines. Not
CC detected in mesenchymal and epithelial cell lines.
CC -!- DISEASE: Participates in a t(X;18)(p11.2;q11.2) chromosomal
CC translocation that is specifically found in more than 80% of
CC synovial sarcoma and produces the SSX1-SSX1 or SSX1-SSX2 fusion
CC products. These hybrid proteins are probably responsible for
CC transforming activity. Heterogeneity in the position of the
CC breakpoint can occur (low frequency).
CC -!- SIMILARITY: Belongs to the SSX family.
CC -!- SIMILARITY: Contains 1 KRAB-related domain.
[6]
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
[7]
DR EMBL; X86175; CAA60111.1; -
DR EMBL; BC007343; AAH07343.1; -
DR EMBL; BC016957; AAH16957.1; -
DR EMBL; S79332; AAB35379.1; -
DR EMBL; X79200; -; NOT ANNOTATED CDS.
DR EMBL; S79894; AAB35674.1; ALT_INIT.
DR PIR; S55058; S55058.
DR Genbank; HGNC:11336; SSX2.

```

```

DR MIM; 300192; --
DR GO; GO:0006355; P:regulation of transcription, DNA-dependent; NAS.
DR InterPro; IPR001909; KRAB.
DR InterPro; IPR003655; KRAB_related.
DR Pfam; PF01352; KRAB; 1.
DR SMART; SM00349; KRAB; 1.
DR PROSITE; PSS0806; KRAB RELATED; 1.
KW Chromosomal translocation; Proto-oncogene; Multigene family;
KW Transcription regulation.
FT DOMAIN 20 83
FT SITE 68 69
FT SITE 110 111
FT SITE 169 169
FT CONFLICT 169 169
SQ SEQUENCE 188 AA; 21620 MW; BF5D18AA5F45B1B1 CRC64;

Alignment Scores:
Pred. No.: 3,56e-71 Length: 188
Score: 798.00 Matches: 150
Percent Similarity: 86.17% Conservative: 12
Best Local Similarity: 79.79% Mismatches: 26
Query Match: 77.85% Indels: 0
DB: Gaps: 1

US-09-975-856-1 (1-576) x SSX2 HUMAN (1-188)
QY 1 ATGAACGGAGACGACGCTTTGCAAGGAGAGCCAGGAGTATGCTCAATATCAGAGAAG 60
Db 1 MetAsnGlyAspAlaPheAlaArgProThrValGlyAlaGlnIleProGluLys 20
QY 61 TTACGAAAGCCCTTCGATGATATGCAATATCTCTTAAGAAAGATGGGAAAAGATG 120
Db 1 IleGlnLysAlaPheAspAlaLysTyrPheSerLysGluLysMet 40
QY 121 AAATCCTCGAGAAATCGCTATGCTATATGATCAAGCTAACTATGAGTCTAGCTAA 180
Db 1 LysAlaSerGluLysIlePheTyrValTyrMetLysArgLysTyrGluAlaMetThrLys 60
QY 181 CTAGTTTCAAGTCCACCTCCACCTTCATCGTGTAGTAACGGGCTGCAGACTTCCAC 240
Db 1 LeuGlyPheLysAlaThrLeuProPheMetCysAsnLysArgAlaGluAspPheGln 80
QY 241 GGGAAATCAITTTGGTACGATCAACACCAAGCAAGTTCAGTTCCTCAGATGACT 300
Db 1 GlyAsnAspLeuAspAsnAspProAsnArgGlyAsnGlnValGluArgProGlnMetThr 100
QY 301 TTCCGAGCCTCCAGAGAAATCTTCCGAGATCATGCCCAGAGCCAGCAGAGGAAGAA 360
Db 1 PheGlyArgLeuGlnGlyIleSerProLysIleMetProLysLysProAlaGluGly 120
QY 361 AATGGTTGAGGAAGTGCAGAGGATCTGCCCCCAAAATGATCGGAAACAGCTGTGC 420
Db 1 AsnAspSerGluGluValProGluAlaSerGlyProGlnAsnAspGlyLysGluLeuCys 140
QY 421 CCCCCGGGAAATCCAGTACCTTGGAGAGATTAACAGACATCTCGACCCAAAGGGG 480
Db 1 ProProGlyLysProThrThrSerGluLysIleHisGluArgSerGlyProLysArgGly 160
QY 481 AAACATCGCTGGACCCACACACTGCGTGTAGAGAAAGCAGCTGGTGTATTAAGAGATC 540
Db 1 GluHisAlaThrThrHisArgLeuArgGluArgLysGlnLeuValIleTyrGluGluIle 180
QY 541 AGGACCTCGAGGAGATGACGAG 564
Db 181 SerAspProGluGluAspAspGlu 188

RESULT 4
SSX1_HUMAN
ID SSX1_HUMAN STANDARD; PRT; 188 AA.
AC Q16384;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)

```

```

DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE SSX1 protein (Synovial sarcoma, X breakpoint 1).
GN OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Fibrosarcoma;
RX MEDLINE=95292974; PubMed=7539744;
RA Crew A.J., Clark J., Fisher C., Gall S., Grimer R., Chand A.,
RA Shipley J., Gusterson B.A., Cooper C.S.;
RT "Fusion of SYT to two genes, SSX1 and SSX2, encoding proteins with
RT homology to the Kruppel-associated box in human synovial sarcoma.";
RL EMBO J. 14:2333-2340(1995).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Placenta;
RX MEDLINE=22388257; PubMed=12477932;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.P., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaby S.J.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalón D.K., Muzny D.E., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahey J., Helton E., Kettman M., Madan A., Rodrigues S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green B.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smalhus D.E.,
RA Scherch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length
RT human and mouse cDNA sequences.";
RN [3]
RP Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN [3]
RP SEQUENCE OF 111-188 FROM N.A.
RC TISSUE=Synovial sarcoma;
RX MEDLINE=95384157; PubMed=7655467;
RA de Leeuw B., Bailemans M., Olde Weghuis D., Geurts van Kessel A.;
RT "Identification of two alternative fusion genes, SYT-SSX1 and
RT SYT-SSX2, in t(X;18)(p11.2;q11.2)-positive synovial sarcomas.";
RL Hum Mol Genet. 4:1097-1099(1995).
CC -!- FUNCTION: Could act as a modulator of transcription.
CC -!- TISSUE SPECIFICITY: Expressed at high level in the testis.
CC Expressed at low level in thyroid. Not detected in tonsil, colon,
CC lung, spleen, prostate, kidney, striated and smooth muscles.
CC Detected in rhabdomyosarcoma and fibrosarcoma cell lines. Not
CC detected in mesenchymal and epithelial cell lines.
CC -!- DISEASE: Participates in a t(X;18)(p11.2;q11.2) chromosomal
CC translocation that is specifically found in more than 80% of
CC synovial sarcoma and produces the SSX1-SSX1 or SSX1-SSX2 fusion
CC products. These hybrid proteins are probably responsible for
CC transforming activity. Heterogeneity in the position of the
CC breakpoint can occur (low frequency).
CC -!- SIMILARITY: Belongs to the SSX family.
CC -!- SIMILARITY: Contains 1 KRAB-related domain.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; X86174; CAA60110.1; --
CC EMBL; BC001003; AAH01003.1; --

```

```

DR EMBL; S79325; AAB35378.1; -.
DR PIR; S55057; S55057.
DR Genew; HGNC:11335; SSX1.
DR MIM; 312820; -.
DR GO; GO:0005634; C:nucleus; TAS.
DR GO; GO:0003714; F:transcription co-repressor activity; TAS.
DR InterPro; IPR001909; KRAB_related.
DR Pfam; PF01352; KRAB; 1.
DR SMART; SM00349; KRAB; 1.
DR PROSITE; PS50806; KRAB RELATED; 1.
DR Chromosomal translocation; Proto-oncogene; Multigene family;
KW Transcription regulation..
FT DOMAIN 20 83 KRAB-RELATED.
FT SITE 62 63 BREAKPOINT FOR TRANSLATION TO FORM THE
FT SITE 110 111 SSXT-SSX1 FUSION PROTEIN (RARE).
FT SITE 110 111 BREAKPOINT FOR TRANSLATION TO FORM THE
FT SITE 110 111 SSXT-SSX1 FUSION PROTEIN.
SQ SEQUENCE 188 AA; 21931 MW; E440D1B2AE3AE9F7 CRC64;

Alignment Scores:
Pred. No.: 8,966-70 Length: 188
Score: 784.00 Matches: 149
Percent Similarity: 84.04% Conservative: 9
Best Local Similarity: 79.26% Mismatches: 30
Query Match: 76.43% Indels: 0
DB: 1 Gaps: 0

US-09-975-856-1 (1-576) x SSX1_HUMAN (1-188)
QY 1 ATGAACGAGAGCGCGCTTTTCAGAGGAGACCCAGGATGATGCTCAATATCAGAGAG 60
Dd 1 MetanGlyAspAspThrPheAlaLysArgProArgAspAlaLysAlaSerGluLys 20
QY 61 TTACGAAAGCGCTTCGATGATATTCCTCAAGAAAGAGTGGGAAAGATG 120
Dd 21 ArgSerLysAlaPheAspAlaThrTyrPheSerLysLysGluTyrLysMet 40
QY 121 AAATCTCGGAGAAATCGTCTATGTATGAGAGCTAACTATGAGTCACTGATAA 180
Dd 41 LysTyrSerGluLysLysSerTyrValTyrMetLysArgAsnTyrLysAlaMetThrLys 60
QY 181 CTAGTTTCAAGGTCACCTCCACCTTTTCATGCTGCTAGTCAAGGCTGCAGTCCAC 240
Dd 61 LeuGlyPheLysValThrLeuProPheMetCysAsnLysGlnAlaThrAspPheGln 80
QY 241 GGGATGATTTTGGTACGATCGCAACACACAGGAATCAGGTGACGTCCTCAGATGACT 300
Dd 81 GlyAsnAspPheAspAsnAspHisAsnArgArgLysGlnValGluHisProGlnMetThr 100
QY 301 TTCGGCAGCTCCAGAGATCTTCCGAGATCATGCCAGAGAGCCAGCAGAGAGAA 360
Dd 101 PheGlyArgLeuHisArgLysLysLysLysLysLysLysLysLysLysLysLys 120
QY 361 AATGGTTTGAAGGAGTCCAGAGGACATCTGGCCACAAATGATGGAAACAGCTGTC 420
Dd 121 AsnAspSerLysGlyValSerGluAlaSerGlyProGlnAsnArgLysLysLys 140
QY 421 CCCCGGGAATCCAAAGTACCTTGGAGAGATTAACAGACATCTGGACCCCAAGGGGG 480
Dd 141 ProProGlyLysAlaAsnLysSerGluLysLysLysLysLysLysLysLysLys 160
QY 481 AAACATGCTGGACCCAGAGCTGGTGGAGAGAGAGAGAGCTGGTGGTTTATGAGAGATC 540
Dd 161 LysHisAlaThrThrHisArgLeuArgGluArgLysGlnLeuValIleTyrGluGlu 180
QY 541 AGCGACCTGAGGAGAGTACGAGAG 564
Dd 181 SerAspProGluGluAspAspGlu 188

RESULT 5
SSX5_HUMAN
ID SSX5_HUMAN STANDARD; PRT; 188 AA.

```

```

AC O60225; Q96AM3;
DT 15-JUL-1999 (Rel. 38, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE SSX5 protein.
GN SSX5.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORM 1).
RX MEDLINE=98021352; PubMed=9378559;
RA Gure A.O., Tuereci O., Sahin U., Tsang S., Scanlan M.J., Jager E.,
RA Knuth A., Pfeundschnig M., Old L.J., Chen Y.-T.;
RT "SSX: a multigene family with several members transcribed in normal
RT testis and human cancer.";
RL Int. J. Cancer 72:965-971(1997).
RN [2]
RP SEQUENCE FROM N.A. (ISOFORM 2).
RC TISSUE=Skin;
RX MEDLINE=22380257; PubMed=12477932;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Udén T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullany S.J.,
RA Bosak S.A., McSwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahy J., Helton E., Kettman M., Madan A., Rodriguez S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smallus D.E.,
RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length
RT human and mouse cDNA sequences.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
CC -!- FUNCTION: Could act as a modulator of transcription.
CC -!- ALTERNATIVE PRODUCTS:
CC Event=Alternative splicing; Named isoforms=2;
CC Name=1;
CC IsoId=O60225-1; Sequence=Displayed;
CC Name=2;
CC IsoId=O60225-2; Sequence=VSP_006274;
CC Note=No experimental confirmation available;
CC -!- SIMILARITY: Belongs to the SSX family.
CC -!- SIMILARITY: Contains 1 KRAB-related domain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; U0842; AAC05821.1; -.
CC EMBL; BC016640; AAH16640.1; -.
CC Genew; HGNC:11339; SSX5.
CC MIM; 300327; -.
CC InterPro; IPR001909; KRAB.
CC InterPro; IPR003655; KRAB-related.
CC Pfam; PF01352; KRAB; 1.
CC SMART; SM00349; KRAB; 1.
CC PROSITE; PS50806; KRAB RELATED; 1.
CC Multigene family; Transcription regulation; Alternative splicing.
FT DOMAIN 20 83 KRAB-RELATED.
FT VANSPPLIC 23 23 K -> KHPMRQVCGRGHLVNLSPFWKVGREPASSIKALIC

```



```

Db 36 LysThrLeuTyrAsnAspValMetGlnAspIleTyrGluThrValIleSerLeuGlyLeu 55
QY 190 AAGGTACACCTCCACCTTTCATGCGTAGTAAACGGGCTGCAGACTTCACGGGAATGAT 249
Db 56 LysLeuLys-----AsnAsp 60
QY 250 TTGTGTAACGATCGA----- 264
Db 61 ThrGlyAsnAspHisProIleSerValSerThrSerGluIleGlnThrSerGlyCysGlu 80
QY 265 ---AACCACAGCAATCAGGTGCACTGCTCAGATGATCTTCGGC----- 306
Db 81 ValSerLysLysThrArgMetLysIleAlaGlnLysThrMetGlyArgGluAsnProGly 100
QY 307 -----AGCTCCAG-----AGAACTTCCGAGATCATGCCCAAGAGCCA 348
Db 101 AspThrHisSerValGlnLysTrpHisArgAlaPheProArgLysLysArgLysLysPro 120
QY 349 GCAGAGGAGAAATGTTGTGAGGAGTGCAGAGGATCTGGCCCAAAATGATGGG 408
Db 121 AlaThrCysLysGlnGluLeuProLysLeuMetAspLeuHisGlyLysGlyProThrGly 140
QY 409 AAACAG 414
Db 141 GluLys 142

RESULT 7
ID 2398 HUMAN STANDARD; PRT; 642 AA.
AC Q8TD17; Q8TD18; Q9P2K7; Q9UDV8;
DT 28-FEB-2003 (Rel. 41, Created)
DE 10-OCT-2003 (Rel. 42, Last annotation update)
DE Zinc finger protein 398 (Zinc finger DNA binding protein p52/p71).
GN ZNF398 OR ZER6 OR KIAA1339.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI TaxID:9606;
[1]
RP SEQUENCE FROM N.A. (ISOFORMS 1 AND 2).
RX MEDLINE=21883946; PubMed=11779858;
RA Conroy A.T., Sharma M., Holcz A.E., Wu C., Sun Z., Weigel R.J.;
RT "A novel zinc finger transcription factor with two isoforms that are
RL differentially repressed by estrogen receptor-alpha."
RN J. Biol. Chem. 277:9326-9334(2002).
[2]
RP SEQUENCE FROM N.A.
RA Bemis G., Langston Y., Tucci S.;
RL Submitted (JUN-1998) to the EMBL/GenBank/DBJ databases.
[3]
RP SEQUENCE FROM N.A. (ISOFORM 1).
RX MEDLINE=22388257; PubMed=12477932;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins P.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Jordan B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Zordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong F.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Udell T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullany S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahey J., Helton E., Kettelman M., Madan A., Rodrigues S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smalhus D.E.,
RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length
human and mouse cDNA sequences."

```

---

```

RL 36 LysThrLeuTyrAsnAspValMetGlnAspIleTyrGluThrValIleSerLeuGlyLeu 55
RN SEQUENCE OF 234-642 FROM N.A.
RX MEDLINE=20181126; PubMed=10718198;
RA Nagase T., Kikuno R., Ishikawa K.-I., Hirose M., Ohara O.;
RT "Prediction of the coding sequences of unidentified human genes. XVI.
RT The complete sequences of 150 new cDNA clones from brain which code
RL for large proteins in vitro."
RL DNA Res. 7:65-73(2000).
CC -!- FUNCTION: Function as a transcriptional activator.
CC -!- SUBCELLULAR LOCATION: Nuclear (Probable).
CC -!- ALTERNATIVE PRODUCTS:
CC Event-Alternative splicing; Named isoforms=2;
CC Name=1; Synonyms=p71;
CC IsoId=Q8TD17-1; Sequence=Displayed;
CC Name=2; Synonyms=p52;
CC IsoId=Q8TD17-2; Sequence=VSP 006926;
CC -!- INDUCTION: By estrogen receptor alpha.
CC -!- SIMILARITY: BELONGS TO THE KRUEPPEL FAMILY OF C2H2-TYPE ZINC-
CC FINGER PROTEINS.
CC -!- SIMILARITY: Contains 1 KRAB domain.
CC -!- CAUTION: Ref.2 sequence differs from that shown due to erroneous
CC gene model prediction.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AY049744; AAK32789.1; -
DR EMBL; AY049743; AAK32788.1; -
DR EMBL; AC004890; AAD45824.1; ALT_SEQ.
DR EMBL; BC043295; AAH43295.1; -
DR EMBL; AB037760; BAA92577.1; -
DR TRANSFAC; T05129; -
DR Genew; HGNC:18373; ZNF398.
DR GO; GO:0005634; C:nucleus; NAS.
DR GO; GO:0016563; F:transcriptional activator activity; NAS.
DR GO; GO:0006355; P:regulation of transcription, DNA-dependent; NAS.
DR InterPro; IPR001309; KRAB.
DR InterPro; IPR007087; Znf_C2H2.
DR Pfam; PF01352; KRAB; 1.
DR Pfam; PF00096; zf-C2H2; 8.
DR ProDom; PD000003; Znf_C2H2; 2.
DR SMART; SM00349; KRAB; 1.
DR SMART; SM00355; Znf_C2H2; 8.
DR PROSITE; PS00805; KRAB; 1.
DR PROSITE; PS00028; ZINC_FINGER_C2H2_1; 7.
DR PROSITE; PS0157; ZINC_FINGER_C2H2_2; 8.
KW Transcription regulation; Activator; DNA-binding; Zinc-finger;
KW Metal-binding; Nuclear protein; Repeat; Alternative splicing.
FT DOMAIN 143 214
FT ZN_FING 343 364 C2H2-TYPE (ATYPICAL).
FT ZN_FING 370 392 C2H2-TYPE (DEGENERATE).
FT ZN_FING 398 420 C2H2-TYPE.
FT ZN_FING 427 449 C2H2-TYPE.
FT ZN_FING 455 477 C2H2-TYPE.
FT ZN_FING 483 505 C2H2-TYPE.
FT ZN_FING 511 533 C2H2-TYPE.
FT ZN_FING 539 561 C2H2-TYPE.
FT ZN_FING 567 590 C2H2-TYPE.
FT VARSPLOC 1 171 Missing (in isoform 2).
SQ SEQUENCE 642 AA; 71311 MW; 69AA38FC84FF633 CRC64;
/FTId=VSP 006926.
Alignment Scores:
Pred. No.: 0.0593 Length: 642
Score: 101.00 Matches: 31
Percent Similarity: 41.60% Conservative: 21
Best Local Similarity: 24.80% Mismatches: 41

```



```

Db 139 SerLeuLeuMetGluAspIlePheGlyysGluThrProSerGlyValThrMetGluArg 158
QY 361 AATGGTTTGAAGGAA 375
Db 159 AlaGlyLeuGlyGlu 163

RESULT 9
PK3G_RAT
ID PK3G_RAT STANDARD; PRT; 1505 AA.
AC O70173;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DE Phosphatidylinositol-4-phosphate 3-kinase C2 domain-containing gamma
DE polypeptide (EC 2.7.1.154) (Phosphoinositide 3-Kinase-C2-gamma)
DE (PtdIns-3-kinase C2 gamma) (PI3K-C2gamma).
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A., FUNCTION, TISSUE SPECIFICITY, AND DEVELOPMENTAL
RP STAGE.
RC TISSUE=Regenerating liver;
RX MEDLINE=98184888; PubMed=9516481;
RA Ono F., Nakagawa T., Saito S., Owada Y., Sakagami H., Goto K.,
RA Suzuki M., Matsuno S., Kondo H.;
RT "A novel class II phosphoinositide 3-kinase predominantly expressed in
RT the liver and its enhanced expression during liver regeneration.";
RL J. Biol. Chem. 273:7731-7736 (1998).
CC -!- FUNCTION: IN VITRO, PHOSPHORYLATES PTDINS AND PTDINS4P BUT NOT
CC PTDINS(4,5)P2.
CC -!- CATALYTIC ACTIVITY: ATP + 1-phosphatidyl-1D-myo-inositol 4-
CC phosphate = ADP + 1-phosphatidyl-1D-myo-inositol 3,4-bisphosphate.
CC -!- SUBCELLULAR LOCATION: Membrane-associated (By similarity).
CC -!- TISSUE SPECIFICITY: PREDOMINANTLY EXPRESSED IN NORMAL LIVER. HIGH
CC LEVELS ALSO FOUND IN REGENERATING LIVER. VERY LOW LEVELS FOUND IN
CC HEART AND TESTIS.
CC -!- DEVELOPMENTAL STAGE: HIGHER LEVELS OF EXPRESSION FOUND IN ADULT
CC LIVER THAN IN FETAL LIVER.
CC -!- SIMILARITY: Belongs to the PI3/P14-kinase family.
CC -!- SIMILARITY: Contains 1 C2 domain.
CC -!- SIMILARITY: Contains 1 pbox homology (PX) domain.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AB009636; BAA25634.1; -.
CC InterPro; IPR008938; ARM.
CC InterPro; IPR000008; C2.
CC InterPro; IPR008973; C2_CaLB.
CC InterPro; IPR000403; PI3_P14_kinase.
CC InterPro; IPR002420; PI3K_C2.
CC InterPro; IPR000341; PI3K_ras_bind.
CC InterPro; IPR001263; PI3Ka.
CC InterPro; IPR001693; PX.
CC Pfam; PF00168; C2; 1.
CC Pfam; PF00454; PI3_P14_Kinase; 1.
CC Pfam; PF00792; PI3K_C2; 1.
CC Pfam; PF00794; PI3K_rbd; 1.
CC Pfam; PF00613; PI3Ka; 1.
CC Pfam; PF00787; PX; 1.
CC SMART; SM00239; C2; 2.
CC SMART; SM00142; PI3K_C2; 1.
CC SMART; SM00144; PI3K_rbd; 1.
CC SMART; SM00145; PI3Ka; 1.

```

```

DR SMART; SM00146; PI3Kc; 1.
DR SMART; SM00312; PX; 1.
DR PROSITE; PS00499; C2_DOMAIN_1; FALSE_NEG.
DR PROSITE; PS00004; C2_DOMAIN_2; FALSE_NEG.
DR PROSITE; PS00915; PI3_4_KINASE_1; 1.
DR PROSITE; PS00916; PI3_4_KINASE_2; 1.
DR PROSITE; PS00290; PI3_4_KINASE_3; 1.
DR PROSITE; PS00195; PX; 1.
KW Transferase; Kinase; Membrane; Multigene family.
FT DOMAIN 976 1240 PI3K/PI4K.
FT DOMAIN 1259 1371 PX.
FT DOMAIN 1402 1499 C2_DOMAIN.
SQ SEQUENCE 1505 AA; 170974 MW; 5ED4C2239968C4B2 CRC64;

Alignment Scores:
Pred. No.: 0.95 Length: 1505
Score: 89.50 Matches: 33
Percent Similarity: 42.24% Conservative: 16
Best local Similarity: 28.45% Mismatches: 50
Query Match: 8.89% Indels: 17
DB: 1 Gaps: 5

US-09-975-856-1 (1-576) x PK3G_RAT (1-1505)
QY 490 AGGCATGTTTCCCTTTGGGTCCAGATGCTCTTAACTCTTCCAGAGTACTTGGAT 431
Db 83 ArgHisPheAsnGluPheThrSerGln-----SerProHisPheSerGln 97
QY 430 TTCGGGGGGGCACAGCGTTTCCCATCATTTTGTGGGCCAGATCGCTCTGGCACTTCTCT 371
Db 98 LeuProPheGlyLysAlaSerAla-----IleGlyPheAsnProAlaValLeuPro 114
QY 370 TCAACACATTTTCTCTCTCTGCTGCTTCTTGGGCATGATCTTCGGGAAGATCTCTCGA 311
Db 115 AlahisGlnPheIleHisGluGlyAlaSerTrpArgAsnProThrArgLysTyrHisGly 134
QY 310 GCCTGCCGAAAGTCATCTGAGGACGCTTCAACTGATTCCTGTGGTTTGCATCGT---TAC 254
Db 135 GlyGluAspProArgPheSerAlaLeuThrProSerSerThrGlyLeuAspLysCysHis 154
QY 253 CAAATCATCTCCGTGGGAAGCTCGAGCCGCTTACTAGCATGAAAGTGGAGGTGA 194
Db 155 GlnGlnGlnSerGlyThrGluHisCysAsnTyrTyrVal-----Glu 169
QY 193 CTTGGAACCTAGTGTAGTACCTCATGACCTCATAGTTAGTCTCATATACA 146
Db 170 ProGluAsnAsnVal-----ProHisTyrSerProTyrSer 182

RESULT 10
FM01_CANFA
ID FM01_CANFA STANDARD; PRT; 531 AA.
AC Q951A2;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Dimethylalanine monooxygenase [N-oxide forming] 1 (EC 1.14.13.8)
DE (Hepatic flavin-containing monooxygenase 1) (FM01) (Dimethylalanine
DE oxidase 1).
OS Canis familiaris (Dog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.
OX NCBI_TaxID=9615;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Liver;
RA Lattard V., Longin-Sauvageon C., Lachuer J., Buronfosse T., Benoit B.;
RT "Cloning, sequencing and tissue dependent expression of FM01 and FM03
RT in the dog.";
RL Submitted (MAY-2001) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: This protein is involved in the oxidative metabolism of
CC a variety of xenobiotics such as drugs and pesticides.
CC -!- CATALYTIC ACTIVITY: N,N-dimethylaniline + NADPH + O(2) = N,N-

```



```

CC dimethylalanine N-oxide + NADP(+) + H(2)O.
CC -I- COFACTOR: FAD (By similarity).
CC -I- SUBCELLULAR LOCATION: Microsomal.
CC -I- TISSUE SPECIFICITY: Liver.
CC -I- SIMILARITY: Belongs to the FMO family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AF384053; AAK97433.1; .
DR InterPro: IPR000759; Admrdx_reductase.
DR InterPro: IPR001327; FAD_PyR_redox.
DR Pfam: PR00743; FMO-like; 1.
DR PRINTS: PR00419; ADXRUTASE.
DR PRINTS; PR00368; FADPMR.
DR PRINTS; PR00370; FMOXYGENASE.
KW Oxidoreductase; Monooxygenase; NADP; Flavoprotein; FAD; Microsome;
KW Transmembrane; Multigene family; Acetylation.
FT INIT_MET 0 0 BY SIMILARITY.
FT MOD_RES 1 1 ACETYLATION (BY SIMILARITY).
FT NP_BIND 8 13 FAD (ADP PART) (POTENTIAL).
FT NP_BIND 190 195 NADP (POTENTIAL).
SQ SEQUENCE 531 AA; 59927 MW; 9F3458484540521F CRC64;
Alignment Scores:
Pred. No.: 0.996 Length: 531
Score: 88.50 Matches: 33
Percent Similarity: 38.21% Conservative: 14
Best Local Similarity: 26.83% Mismatches: 36
Query Match: 8.79% Indels: 40
DB: 1 Gaps: 4
US-09-975-856-1 (1-576) x FMO1_CANFA (1-531)
QY 565 ACTCGTCATCTCTCCAGGTCGCTGATCTCTTCATAAACCACGAGCTGCTTCTCTCAC 506
Db 43 ThrGluHisValGluGluGlyArgAlaSerLeuTyrlsSerVal----- 57
QY 505 GCAGTCGTGGGTCACGAGCATGTTCCCTTTGGGTCCAGATGCTTGTAACTTCT 446
Db 58 -----ValSerAsnSerCysLysGluMetSerCys----- 67
QY 445 CCAAGGTACTTGATTTCCCGGGGGGCACAGCTGTTTCCCATCATTTTGGGCCAGATG 386
Db 68 -----TyrSerAspPhePro-----PheProGluAspTyr 77
QY 385 CCTCTGGCACTTCTCTCAACCACTTCTTCCCTCTGCTGCTTCTTGGGCATGATCTTCG 326
Db 78 ProAsnTyrValProAsnSerGlnPheLeuGluTyrLeuLysMetTyrAla-Asn----- 95
QY 325 GGAAGATTCTCTGGAGGCTGCCAAGATCATCTGACGAGCTTCAACTGATCTCTGTGT 266
Db 96 -----ArgPheSerLeuLeuLysCys11 103
QY 265 TTCGATCGTTTACCAAAATCATCTCCCGTGAAGTCTGCAGCCGCTTTTACTACGCATGAAG 206
Db 103 eArgPheLysThrLysValCysLysValThrLysCysProAspPheThrValThrGlyG 123
QY 205 GTGGGAG 199
Db 123 nTyrGlu 125
RESULT 11
REPL_MOUSE
ID REPL_MOUSE STANDARD; PRT; 743 AA.
AC O54916; Q8C9J9; Q99LR8;
DT 10-OCT-2003 (Rel. 42, Created)

```

```

DT 10-OCT-2003 (Rel. 42, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE RalBP1 associated Eps domain containing protein 1 (RalBP1-interacting
DE protein 1).
DE REPS1.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]_TaxID=10090;
RP SEQUENCE FROM N.A. (ISOFORM 1), AND CHARACTERIZATION.
RC TISSUE=Muscle;
RX MEDLINE=98058900; PubMed=9395447;
RA Yamaguchi A., Urano T., Goi T., Feig L.A.;
RT "An eps homology (EH) domain protein that binds to the ral-GTPase
RL J. Biol. Chem. 272:31230-31234 (1997).
RN [2]
RP SEQUENCE OF 9-743 FROM N.A. (ISOFORM 2).
RC STRAIN=C57BL/6J; TISSUE=Thymus;
RX MEDLINE=22354683; PubMed=12466851;
RA Okazaki Y., Furuno M., Kasukawa T., Adachi J., Bono H., Kondo S.,
RA Nikaide I., Oato N., Saito R., Suzuki H., Yamanaka I., Kiyosawa H.,
RA Baldarelli R., Hill D.P., Bult C., Hume D.A., Quackenbush J.,
RA Schriml L.M., Kanapin A., Matsuda H., Batalov S., Beisel K.W.,
RA Blake J.A., Bradt D., Brusic V., Chothia C., Corbani L.E., Cousins S.,
RA Dalla E., Dragani T.A., Fletcher C.F., Forrest A., Frazer K.S.,
RA Gaasterland T., Gariboldi M., Gissi C., Godzik A., Gough J.,
RA Grimmond S., Gustincich S., Hirokawa N., Jackson I.J., Jarvis E.D.,
RA Kanai A., Kawaji H., Kawasawa Y., Kedzierski R.M., King B.L.,
RA Konagaya A., Kurochkin I.V., Lee Y., Lenhard B., Lyons P.A.,
RA Maglott D.R., Maltais L., Marchionni L., McKenzie L., Miki H.,
RA Nagashima T., Numata K., Okido T., Pavan W.J., Pertea G., Pesole G.,
RA Petrovsky N., Pillai R., Pontius J.U., Qi D., Ramachandran S.,
RA Ravasi T., Reed J.C., Reed D.J., Reid J., Ring B.Z., Ringwald M.,
RA Sandelin A., Schneider C., Sempile C.A., Setou M., Shimada K.,
RA Sultana R., Takenaka Y., Taylor M.S., Teasdale R.D., Tomita M.,
RA Verardo R., Wagner L., Wahlstedt C., Wang Y., Watanabe Y., Wells C.,
RA Wilming L.G., Wynshaw-Boris A., Yanagisawa M., Yang I., Yang L.,
RA Yuan Z., Zavalon M., Zhu Y., Zimmer A., Carninci P., Hayatsu N.,
RA Hirozane-Kishikawa T., Konno H., Nakamura M., Sakazume N., Sato K.,
RA Shiraki T., Waki K., Kawai J., Aizawa K., Arakawa T., Fukuda S.,
RA Hara A., Hashizume W., Imotani K., Ishii Y., Itoh M., Kagawa I.,
RA Miyazaki A., Sakai K., Sasaki D., Shibata K., Sainagawa A.,
RA Yasunishi A., Yoshino M., Waterston R., Lander E.S., Rogers J.,
RA Birney E., Hayashizaki Y.;
RT "Analysis of the mouse transcriptome based on functional annotation of
RT 60,770 full-length cDNAs.";
RL Nature 420:563-573 (2002).
RN [3]
RP SEQUENCE OF 447-743 FROM N.A. (ISOFORM 1).
RC TISSUE=Breast tumor;
RX MEDLINE=22388257; PubMed=12477932;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Udwin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,
RA Bosak S.A., McSwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalón D.K., Muzny K.C., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahey J., Helton E., Kettman M., Madan A., Rodrigues S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smalusz D.E.,
RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length
RT human and mouse cDNA sequences.";

```



Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).

[4]  
 RN STRUCTURE BY NMR OF 227-318.  
 RX MEDLINE=21285759; PubMed=11389591;  
 RA Kim S., Cullis D.N., Feig L.A., Balleja J.D.;  
 RT "Solution structure of the Rep1 EH domain and characterization of  
 RT its binding to NPF target sequences."  
 RL Biochemistry 40:6776-6785(2001).  
 CC -!- FUNCTION: May coordinate the cellular actions of activated EGF  
 CC receptors and Ral-GTPases.  
 CC -!- SUBUNIT: Homodimer (Potential). Interacts with RALBP1, CRK and  
 CC GRB2. Binding to RALBP1 does not affect its Ral-binding activity.  
 CC Forms a complex with the SH3 domains of CRK and GRB2 which may  
 CC link it to an EGF-responsive tyrosine kinase.  
 CC -!- ALTERNATIVE PRODUCTS:  
 CC Event=Alternative splicing; Named isoforms=2;  
 CC Name=1;  
 CC IsoId=O54916-1; Sequence=Displayed;  
 CC Name=2;  
 CC IsoId=O54916-2; Sequence=VSP\_007956, VSP\_007957;  
 CC Notes=Due to intron retention. No experimental confirmation  
 CC available;  
 CC -!- TISSUE SPECIFICITY: Expressed in all tissues examined. The highest  
 CC level expression was found in the kidney and testis.  
 CC -!- PTM: EGF stimulates phosphorylation on Tyr-residues.  
 CC -!- SIMILARITY: Contains 1 EH domain.  
 CC -!- SIMILARITY: Contains 1 EF-hand calcium-binding domain.  
 CC -!- CAUTION: Ref.3 sequence differs from that shown due to a  
 CC frameshift in position 719.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC -----  
 CC EMBL; AF031939; AAB94736.1; -;  
 CC EMBL; AK041967; BAC31117.1; ALT INIT.  
 CC EMBL; BC002256; AAB02256.1; ALT\_FRAME.  
 CC PIR; T09173; T09173.  
 CC PDB; 1RT6; 18-JUL-01.  
 CC MGD; MGI:1196373; Reps1.  
 CC InterPro; IPR002048; EF-hand.  
 CC InterPro; IPR00261; EPS15\_homology.  
 CC Pfam; PF00036; efhand; 1.  
 CC SMART; SM00027; EH; 1.  
 CC PROSITE; PS00018; EF\_HAND; 1.  
 CC PROSITE; PS00031; EH; 1.  
 CC Calcium-binding; Coiled coil; Phosphorylation; Alternative splicing;  
 CC 3D-structure.  
 CC DOMAIN 233 324 EH.  
 CC CA\_BIND 279 290 EF-HAND (POTENTIAL).  
 CC FT DOMAIN 488 551 PRO-RICH.  
 CC FT DOMAIN 599 743 INTERACTION WITH RALBP1.  
 CC FT DOMAIN 692 738 COILED COIL (POTENTIAL).  
 CC FT MOD\_RES 236 236 PHOSPHORYLATION (POTENTIAL).  
 CC FT VARSPPLIC 368 402 OWETFSRSSSQTLPQDSNIAPADPDTAIVHPV ->  
 CC VSKTSLSLLEISLFTGRSFKQDRFTAGVLYQYHTP (in  
 CC isoform 2).  
 CC /FTId=VSP\_007956.  
 CC Missing (in isoform 2).  
 CC /FTId=VSP\_007957.  
 CC FT VARSPPLIC 403 743  
 CC SEQUENCE 743 AA; 80598 MW; 25510D1I254CF4A6 CRC64;  
 CC  
 CC Alignment Scores:  
 CC Pred. No.: 1.18 Length: 743  
 CC Score: 88.00 Matches: 40  
 CC Percent Similarity: 38.92% Conservative: 25  
 CC Best Local Similarity: 23.95% Mismatches: 41  
 CC Query Match: 8.59% Indels: 61  
 CC DB: 1 Gaps: 9

US-09-975-856-1 (1-576) x REPI\_MOUSE (1-743)

QY 91 TACTTCTTAAGAAAGAGTGGGAAAAGATGAATCTCTCGAGAAAATCTCTATGTGTAT 150  
 ||||| : : : : :  
 Db 11 TyrPheGlyArgSerGlnPhe : : : : : Tyr 18  
 : : : : :  
 QY 151 ATGAAGCTAAACTATCAGGTCAATGACTAACTAGGTTTC : : : : : 189  
 ||||| : : : : :  
 Db 19 IleAlaLeuLeuValAlaValAlaGlnSerGlyPheProLeuArgValGluSerIle 38  
 : : : : :  
 QY 190 -----AGGTCACTCCCTCCACCTTTTCATCGCGGTAGTAAA----- 222  
 : : : : :  
 Db 39 AsnThrVallyAspLeuProLeuProArgPheValAlaSerLysAsnGluGlnGluSer 58  
 : : : : :  
 QY 223 CGGCTGCGAGCTTCCCGGGAATCATTTTGTAAAGTCAACACAC----- 270  
 ||||| : : : : :  
 Db 59 ArgLeuAlaAlaSerTy:SerSerAspSerGluAsnGlnGlySerTy:SerGlyValIle 78  
 : : : : :  
 QY 271 -----AGGAATCAGGTTGAACCTCTCAGATGATCTTCGGCAGCTCCAG 315  
 ||||| : : : : :  
 Db 79 ProProProGlyArgGlyGlnVallyLysGlyProGlySerHisAspAlaValGln 98  
 : : : : :  
 QY 316 AGAATCTCCGAAAGATCATGCCAAGAGCCAGCAGAGAGAAAGAAATGTTTGAAGAA 375  
 : : : : :  
 Db 99 -----ProArgProSerAlaGluGlnGln----- 106  
 : : : : :  
 QY 376 GTGCCAGAGGATCTGGCCCAAAAATGATGGAAACAGCTGTGCCCCCGGGAATCCA 435  
 ||||| : : : : :  
 Db 107 -----GluProAlaSerProValValSerProGlnGln---SerProThrSerPro 123  
 : : : : :  
 QY 436 AGTACTCTGGAGAGATTAAACAAGACATCTGACCCCAAGGGGGAACATG----- 487  
 ||||| : : : : :  
 Db 124 HisThr-TrpArgLysHisSerArgHis-----ProSerGlyGlyAsnSerGluArgPr 141  
 : : : : :  
 QY 488 -----CCTGGACCC 496  
 ||||| : : : : :  
 Db 141 oLeuThrGlyProGlyPro 147

RESULT 12  
 FMO1\_PIG  
 ID FMO1\_PIG STANDARD; PRT; 531 AA.  
 AC P16549;  
 DT 01-AUG-1990 (Rel. 15, Created)  
 DT 01-OCT-1993 (Rel. 27, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Dimethylalanine monooxygenase [N-oxide forming] 1 (EC 1.14.13.8)  
 DE (Hepatic flavin-containing monooxygenase 1) (FMO 1) (Dimethylalanine  
 DE oxidase 1).  
 DE FMO1 OR FMO-1.  
 OS Sus scrofa (Pig).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.  
 OX NCBI\_TaxID=9823;  
 RN [1] SEQUENCE FROM N.A., AND SEQUENCE OF 136-150 AND 308-317.  
 RP TISSUE=Liver;  
 RC MEDLINE=90212556; PubMed=2322534;  
 RX Gasser R., Tynes R.E., Lawton M.P., Korsmeyer K.K., Ziegler D.M.,  
 RA Philpot R.M.;  
 RT "The flavin-containing monooxygenase expressed in pig liver: primary  
 RT sequence, distribution, and evidence for a single gene."  
 RL Biochemistry 29:119-124 (1990).  
 RN [2]  
 RP SEQUENCE OF 1-13 AND 184-201, AND ACETYLATION.  
 RC TISSUE=Liver;  
 RX MEDLINE=90343821; PubMed=2383273;  
 RA Guan S.H., Falick A.M., Cashman J.R.;  
 RT "N-terminus determination: FAD and NADP binding domain mapping of hog  
 RT liver flavin-containing monooxygenase by tandem mass spectrometry."  
 RL Biochem. Biophys. Res. Commun. 170:937-943 (1990).  
 RN [3]  
 RP SEQUENCE OF 185-207.

RC TISSUE=liver;  
 RX MEDLINE=95278229; PubMed=7758472;  
 RA Wu R.-F., Ichikawa Y.;  
 RT "An essential lysyl residue (Lys208) in the substrate-binding site of  
 RL porcine PAD-containing monooxygenase.";  
 RN Eur. J. Biochem. 229:749-753 (1995).  
 [4]  
 RP CARBOHYDRATE-LINKAGE SITE.  
 RX MEDLINE=98451545; PubMed=9778310;  
 RA Korsmeyer K.K., Guan S., Yang Z.C., Falick A.M., Ziegler D.M.,  
 RT Cashman J.R.;  
 RT "N-glycosylation of pig flavin-containing monooxygenase form 1:  
 RT determination of the site of protein modification by mass  
 RT spectrometry.";  
 RL Chem. Res. Toxicol. 11:1145-1153 (1998).  
 CC -!- FUNCTION: This protein is involved in the oxidative metabolism of  
 CC a variety of xenobiotics such as drugs and pesticides.  
 CC -!- CATALYTIC ACTIVITY: N,N-dimethylaniline + NADPH + O(2) = N,N-  
 CC dimethylaniline N-oxide + NADP(+) + H(2)O.  
 CC -!- COFACTOR: FAD.  
 CC -!- SUBCELLULAR LOCATION: Microsomal.  
 CC -!- TISSUE SPECIFICITY: Liver.  
 CC -!- SIMILARITY: Belongs to the FMO family.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 DR EMBL; M32031; AAA31033.1; -.  
 DR PIR; A33768; A33768.  
 DR InterPro; IPR000759; Adrxndx\_reductase.  
 DR InterPro; IPR001327; FAD\_pyf\_redox.  
 DR InterPro; IPR000960; Flav\_cont\_mnoxgn.  
 DR Pfam; PF00743; FMO-like; 1.  
 DR PRINTS; PR00419; ADXRDXTASE.  
 DR PRINTS; PR00368; FADPNR.  
 DR PRINTS; PR00370; FMOXYGENASE.  
 KW Oxidoreductase; Monooxygenase; NADP; Flavoprotein; FAD; Microsome;  
 KW Transmembrane; Multigene family; Acetylation; Glycoprotein.  
 FT INIT\_MET 0 0  
 FT MOD\_RES 1 1  
 FT NP\_BIND 8 13  
 FT NP\_BIND 190 195  
 FT ACT\_SITE 207 207  
 FT CARBOHYD 119 119  
 FT SEQUENCE 531 AA; 59821 MW; 5E475E6BD81C8157 CRC64;  
 Alignment Scores:  
 Pred. No.: 1.25 Length: 531  
 Score: 87.50 Matches: 33  
 Percent Similarity: 37.40% Conservative: 13  
 Best Local Similarity: 26.83% Mismatches: 37  
 Query Match: 8.69% Indels: 40  
 DB: 1 Gaps: 4  
 US-09-975-856-1 (1-576) x FMO1\_PIG (1-531)  
 QY 565 ACTCGTCATCTTCTCAGGTCGCTGATCTCTTCATAAACACACAGCTGCTTCTCTCAC 506  
 Db 43 ThrGluHisValGluGluGlyArgAlaSerLeuTyrLysSerVal----- 57  
 QY 505 GCAGTCTGTGGTCCAGGCATGTTTCCCTTTTGGGTCCAGATGCTCTGTTAATCTTCT 446  
 Db 58 -----ValSerAsnSerCysLysLysLysMetSerCys----- 67  
 QY 445 CCAAGGTACTGTGATTTCCGGGGGGGCACAGCTGTTTCCCATCATTTTGTGGGCCAGATG 386  
 Db 68 -----TyrProAspPhePro-----PheProGluAspTyr 77

QY 385 CCTTGGCACATCTCTTCAACCATTTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCG 326  
 Db 78 ProAsnTyrValProAsnSerHisPheLeuGluTyrLeu----- 90  
 QY 325 GGAAGATTCTCTGGAGGCTGCCGAAGTCACTGAGGAGTTCACACTGATCTCTCTGTT 266  
 Db 91 -----Arg-MetTyrAlaAsnGlnPheAsnLeuLeuLysCys11 103  
 QY 265 TTCGATCGTTTACAAAATCATTTCCCGTGAAGTCTGCAGCCGCTTTTACTACGCATGAAAG 206  
 Db 103 eglNPhelysThrLysValCysSerValThrLysHisGluAspPheAsnThrThrGlyGI 123  
 QY 205 GTGGGAG 199  
 Db 123 nlrpasp 125  
 RESULT 13  
 REPI\_HUMAN  
 ID REPI\_HUMAN STANDARD; PRT; 744 AA.  
 AC Q96D71; Q8NDR7; Q8WU62; Q9BXY9;  
 DT 10-OCT-2003 (Rel. 42, Created)  
 DT 10-OCT-2003 (Rel. 42, Last sequence update)  
 DT 10-OCT-2003 (Rel. 42, Last annotation update)  
 DE RalBP1 associated Eps domain containing protein 1 (RalBP1-interacting  
 DE protein 1).  
 GN REPS1.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A. (ISOFORM 1).  
 RA Mao Y., Xie Y., Zhou Z., Zhao W., Zhao S., Wang W., Huang Y., Wang S.,  
 RA Tang R., Chen X., Wu C.;  
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A. (ISOFORMS 2 AND 3).  
 RC TISSUE=Brain, and Lymph;  
 RX MEDLINE=22388257; PubMed=12477932;  
 RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,  
 RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,  
 RA Altshul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,  
 RA Hopkins R.F., Jordan H., Moore I., Max S.I., Wang J., Hsieh F.,  
 RA Diachenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,  
 RA Stapleton M., Soares M.B., Bonaldo M.P., Casavant T.L., Scheetz T.E.,  
 RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,  
 RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullany S.J.,  
 RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,  
 RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,  
 RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,  
 RA Fahey J., Helton E., Kettman M., Madan A., Rodriguez S., Sanchez A.,  
 RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,  
 RA Alakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,  
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,  
 RA Butlerfield Y.S.N., Krzywinski M.I., Skalska U., Smalhus D.E.,  
 RA Scherch A., Schein J.E., Jones S.J.M., Marra M.A.;  
 RT "Generation and initial analysis of more than 15,000 full-length  
 RT human and mouse cDNA sequences.";  
 RN Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).  
 RN [3]  
 RP SEQUENCE OF 274-743 FROM N.A. (ISOFORM 1).  
 RC TISSUE=Brain;  
 RA Bloesker H., Boecher M., Brandt P., Mewes H.-W., Weil B., Wiemann S.;  
 RL Submitted (JUL-2002) to the EMBL/GenBank/DBJ databases.  
 CC -!- FUNCTION: May coordinate the cellular actions of activated EGF  
 CC receptors and Ral-GTPases (By similarity).  
 CC -!- SUBUNIT: Homodimer (Potential). Interacts with RALBP1, CRK and  
 CC GRB2. Binding to RALBP1 does not affect its Ral-binding activity.  
 CC Forms a complex with the SH3 domains of CRK and GRB2 which may  
 CC link it to an EGF-responsive tyrosine kinase (By similarity).  
 CC -!- ALTERNATIVE PRODUCTS:  
 CC -!- Event=Alternative splicing; Named isoforms=3;  
 CC Name=1;

CC IsoId=Q96D71-1; Sequence=Displayed;  
 CC Name=2;  
 CC IsoId=Q96D71-2; Sequence=VSP\_007953, VSP\_007954;  
 CC Name=3;  
 CC IsoId=Q96D71-3; Sequence=VSP\_007955;  
 CC -!- PTM: EGF stimulates phosphorylation on Tyr-residues (By  
 CC similarity).  
 CC -!- SIMILARITY: Contains 1 EH domain.  
 CC -!- SIMILARITY: Contains 1 EF-hand calcium-binding domain.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL; AF251052; AAK34942.1; -.  
 CC EMBL; BC012764; AAH12764.1; -.  
 CC EMBL; BC021211; AAH21211.1; -.  
 CC EMBL; AL831900; CAD38569.1; -.  
 CC Genew; HGNC:15578; REPS1.  
 CC InterPro; IPR002048; EF-hand.  
 CC InterPro; IPR000261; EPLS5 homology.  
 CC Pfam; PF00036; ehand; 1.  
 CC SMART; SM00027; EH; 1.  
 CC PROSITE; PS00018; EF\_HAND; 1.  
 CC PROSITE; PS00031; EH; 1.  
 CC Calcium-binding; Coiled coil; Phosphorylation; Alternative splicing.  
 KW DOMAIN 233 324 EH.  
 FT CA\_BIND 233 324 EF-HAND (POTENTIAL).  
 FT DOMAIN 489 552 PRO-RICH.  
 FT DOMAIN 600 744 INTERACTION WITH RALBP1 (BY SIMILARITY).  
 FT DOMAIN 693 739 COILED COIL (POTENTIAL).  
 FT MOD\_RES 236 236 PHOSPHORYLATION (POTENTIAL).  
 FT VARSPPLIC 336 394 Missing (in isoform 2).  
 FT VARSPPLIC 458 521 Missing (in isoform 2).  
 FT VARSPPLIC 458 521 Missing (in isoform 2).  
 FT VARSPPLIC 458 521 Missing (in isoform 2).  
 FT CONFLICT 576 576 Missing (in isoform 3).  
 FT CONFLICT 665 665 A -> V (IN REF. 1).  
 FT CONFLICT 742 743 V -> I (IN REF. 3).  
 FT CONFLICT 744 80769 SH -> FP (IN REF. 2; AAH12764).  
 SQ SEQUENCE 744 AA; 80769 MW; 1DPF29711DB2B5E4 CRC64;  
 Alignment Scores:  
 Pred. No.: 1.49 Length: 744  
 Score: 87.00 Matches: 39  
 Percent Similarity: 38.89% Conservative: 24  
 Best Local Similarity: 24.07% Mismatches: 48  
 Query Match: 8.49% Indels: 51  
 DB: 1 Gaps: 9  
 US-09-975-856-1 (1-576) x REPI\_HUMAN (1-744)  
 QY 91 TACTTCTCTAAGAAAGAGTGGGAAAGATGAAATCCTCGGAGAAATCGTCTATGTAT 150  
 Db 11 TyrPheGlyArgSerGlnPhe-----Tyr 18  
 QY 151 ATGAAGCTAACTATGAGTCTATGATCACTAACTAGGTTTC----- 189  
 Db 19 IleAlaLeuLysLeuValAlaValAlaGlnSerGlyPheProLeuArgValGluSerIle 38  
 QY 190 -----AAGTCACCTCCACCTTCATCGCTAGTAAA----- 222  
 Db 39 AsnThrValLysAspLeuProLeuProArgPheValAlaSerLysAsnGluGlnGluSer 58  
 QY 223 CGGGCTGCAGACTCCACGGGATGATTGTTGGTAACGATCGAACACACAGGAATCAGGTT 282  
 Db 59 ArgHisAlaAlaSerTyrSerSerAspSerGluAsnGlnGlySerTyrSerGlyValIle 78

QY 283 GAAAGCTCCT-----CAGATGACTTTCGGAGCCTCCAGAGAAATCTTCCCGAAG 330  
 Db 79 ProProProProGlyArgGlyGlnValLysLysGlySerValSerHis-----AspThr 96  
 QY 331 ATCATGCCCAAGAACCCAGCAGAGAGAGAAATGTTTGAAGGAAGTGCCAGAGGATCT 390  
 Db 97 ValGlnProArgThrSerAlaAspAlaGln-----GluProAla 109  
 QY 391 GSCCCACAAAATGATGGGAACAGCTGTCCTCCCGGGAATCCCAAGTACCTTGGAGAAG 450  
 Db 110 SerProValSerProGlnGln---SerProProThrSerProHisThr-TyrArgly 128  
 QY 451 ATTAACAAGACATCTGGACCCAAAGGGGGAACATG-----CCTGG 492  
 Db 128 eHisSerArgHis-----ProSerGlyAsnSerGluArgProLeuAlaGlyProG 146  
 QY 493 ACCC 496  
 Db 146 yPro 147  
 RESULT 14  
 SX11\_CHICK  
 ID SX11\_CHICK STANDARD; PRT; 396 AA.  
 AC P48435;  
 DT 01-FEB-1996 (Rel. 33, Created)  
 DT 01-FEB-1996 (Rel. 33, Last sequence update)  
 DT 10-OCT-2003 (Rel. 42, Last annotation update)  
 DE Transcription factor SOX-11.  
 GN SOX11.  
 OS Gallus gallus (Chicken).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae;  
 OC Gallus.  
 OX NCBI\_TaxID=9031;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Embryo;  
 RX MEDLINE=95267693; PubMed=7748786;  
 RA Uwanogho D., Rex M., Cartwright E.J., Pearl G., Healy C.,  
 RA Scotting P.J., Sharpe P.T.;  
 RA "Embryonic expression of the chicken Sox2, Sox3 and Sox11 genes  
 RT suggests an interactive role in neuronal development."  
 RL Mech. Dev. 49:23-36(1995).  
 CC -!- FUNCTION: May function as switches in neuronal development.  
 CC -!- SUBCELLULAR LOCATION: Nuclear.  
 CC -!- TISSUE SPECIFICITY: Low level expression is seen in  
 CC undifferentiated proliferating cells of neural epithelium. A  
 CC greater expression is seen in the maturing neurons after they  
 CC leave the neural epithelium. It is also found in the gut  
 CC epithelium and adrenal medulla.  
 CC -!- SIMILARITY: Contains 1 HMG box domain.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL; U12534; AAB09664.1; -.  
 CC PIR; I50707; I50707.  
 CC HSSP; P48436; 1SX9.  
 CC InterPro; IPR000910; HMG\_12\_box.  
 CC Pfam; PF00505; HMG\_box; 1.  
 CC SMART; SM00398; HMG; 1.  
 CC PROSITE; PS01118; HMG\_BOX\_2; 1.  
 KW Transcription regulation; DNA-binding; Nuclear protein.  
 FT DNA\_BIND 49 117 HMG\_BOX.  
 FT DOMAIN 162 165 POLY-ALA.  
 FT DOMAIN 204 212 POLY-GLU.  
 FT DOMAIN 288 294 POLY-PRO.  
 FT DOMAIN 332 335 POLY-ARG.

```
SQ SEQUENCE 396 AA; 43503 MW; 8E4B0A45F78BA833 CRC64;
Alignment Scores:
Pred. No.: 1.49 Length: 396
Score: 86.50 Matches: 45
Percent Similarity: 31.75% Conservative: 22
Best Local Similarity: 21.33% Mismatches: 67
Query Match: 8.44% Indels: 77
DB: 1 Gaps: 9
US-09-975-856-1 (1-576) x SX11_CHICK (1-396)
QY 25 AGGAGACCCAGGATGCTCAATATCAGAGAGTTA-----CGAAAGCCTTCGAT 78
Dd ::::::::::: 396
Db 50 LysArgProMetAsnAlaPheMetValTrpSerIleGluAArgIleMetGlu 69
QY 79 -----GATATGCCAAATCTCTCTTAAGAGAGTGGAA 114
Dd ::::::::::: 114
Db 70 GlnSerProAspMetHisAsnAlaGluIleSerLysArgLeuGlyLysArg---TrpLys 88
QY 115 AAGATGAATCTCGGAGAAATCGTATGTATATGAGCTAAACTATGAGGTCATG 174
Dd ::::::::::: 174
Db 89 MetLeuLysAspSerGluLysIle----- 96
QY 175 ACTAAACTAGGTTTCAGGTCACCTCCACCTTTCATCGCTAGT----- 219
Dd ::::::::::: 219
Db 97 -----ProPheIleArgGluAlaGluArgLeuArg 106
QY 220 ---AAACGGGCTGCAGACTTCCACGGGAATGATTTGGTAACTGCAAGAACACAGGAAT 276
Dd ::::::::::: 276
Db 107 LeuLysHisMetAlaAspTyr-----ProAspTyrLysTyrArg--- 119
QY 277 CAGGTGTAACCTCTCAGATCACTTTCGGAGCCTCCAGAGATCTTCCGAGATCATG 336
Dd ::::::::::: 336
Db 120 -----ProArgLysLys 123
QY 337 CCCAAGAACCCAGCAGAGAGAAATGTTTGAAGAGAGTCCAGAGCATCTGGCCCA 396
Dd ::::::::::: 396
Db 124 ProlLysMetAspProSerAlaLysProAsnAlaGlyGlnSerProGluLysAsnAlaPro 143
QY 397 CAAATGATGGAAACAGCTGTGCCCCCGGGAATCCAACTACTCTGGAGAGATTAAC 456
Dd ::::::::::: 456
Db 144 GlyGlyLysSerLysSerAlaLysSerSerGlyLysLysCysSerLysLeuLysAlaAla 163
QY 457 AAGCATCTGACCAAGGAGGAAACATCGCTGGACCCAC----- 498
Dd ::::::::::: 498
Db 164 AlaAlaSerProProLysProGlyAlaLysAlaAlaProHisGlyAspTyrAlaGlyAsp 183
QY 499 -----AGACTGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 531
Dd ::::::::::: 531
Db 184 GluTyrValPheGlyAlaLeuLysValSerSerLysAlaValLysCysValPheValAsp 203
QY 532 GAAGAGATCAGCAGCCCTCGAGAGAGATGACGAG 564
Dd ::::::::::: 564
Db 204 GluGluGluGluAspGluGluAspGluAspGlu 214
RESULT 15
ZN38_HUMAN
ID ZN38_HUMAN STANDARD; PRT; 446 AA.
AC P17036; P13683; Q9NNX8; Q9NNXJ1; Q9UC15; Q9UC16;
DT 01-APR-1990 (Rel. 14, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Zinc finger protein 38 (Zinc finger protein K0X25) (Zinc finger
protein HF.12) (Zinc finger protein 3) (HF3.1 protein).
GN ZNF38 OR K0X25 OR ZNF3.
OS Homo sapiens (Human).
OC Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Ovary, and Skin;
```

DR EMBL; X07290; CAA30269.1; --  
DR PIR; S00754; S00754.  
DR Genew; HGNC:13104; ZNF38.  
DR Genew; HGNC:13089; ZNF3.  
DR MIM; 601261; --  
DR MIM; 194510; --  
DR GO; GO:0005634; C:nucleus; IC.  
DR GO; GO:0003700; F:transcription factor activity; NAS.  
DR GO; GO:0008270; F:zinc ion binding; NAS.  
DR GO; GO:0045321; P:cell activation; NAS.  
DR GO; GO:0006355; P:regulation of transcription, DNA-dependent; NAS.  
DR InterPro; IPR001909; KRAB.  
DR InterPro; IPR007087; Znf\_C2H2.  
DR InterPro; IPR007086; Znf\_C2H2\_sub.  
DR Pfam; PF01352; KRAB; 1.  
DR Pfam; PF00096; zf-C2H2; 8.  
DR PRINTS; PR00048; ZINCFINGER.  
DR ProDom; PD000003; Znf\_C2H2; 8.  
DR SMART; SM00349; KRAB; 1.  
DR SMART; SM00355; Znf\_C2H2; 8.  
DR PROSITE; PS00028; ZINC\_FINGER\_C2H2\_1; 8.  
DR PROSITE; PS0157; ZINC\_FINGER\_C2H2\_2; 8.  
DR PROSITE; PS50805; KRAB; 1.  
KW Transcription regulation; Activator; Zinc-finger; DNA-binding;  
KW Repeat; Metal-binding; Nuclear protein; Differentiation.  
FT DOMAIN 51 123  
FT ZN\_FING 200 227  
FT ZN\_FING 228 255  
FT ZN\_FING 256 283  
FT ZN\_FING 284 311  
FT ZN\_FING 312 339  
FT ZN\_FING 340 367  
FT ZN\_FING 368 395  
FT ZN\_FING 396 423  
FT ZN\_FING 424 451  
FT CONFLICT 21 21  
FT CONFLICT 252 256  
FT CONFLICT 336 387  
SQ SEQUENCE 446 AA; 50932 MW; 67A6926807304782 CRC64;

## Alignment Scores:

Pred. No.:	1.53	Length:	446
Score:	86.50	Matches:	42
Percent Similarity:	36.98%	Conservative:	29
Best Local Similarity:	21.88%	Mismatches:	50
Query Match:	8.44%	Indels:	71
DB:	1	Gaps:	8

US-09-975-856-1 (1-576) x ZN38\_HUMAN (1-446)

QY	40	GATGCTCAATATCAGAGAAGTTACGAAAGGCC-----TTTCGAT	78
Db	35	AspGluMetLeuAlaAlaLeuLeuLeuAlaLysSerGlnGluLeuValThrPheGlu	54
QY	79	GATATTGCCAAATCTCTTAAGAAGAGTGGGAAAGATGAATCTCTGGAGAAATC	138
Db	55	AspValAlaValThrPheIleArgLysGluTyrLysArgLeuGluProAlaGlnArgAsp	74
QY	139	GTCTATGTGTATATGAGCTA---AACTATGAGTCTATGACTAACTAGTTTCAAGTC	195
Db	75	LeuTyrArgAspValMetLeuGluAsnTyr-----	84
QY	196	ACCTCCACCTTTCATCGTAGTAACGGGCTGCAGACTTCCACGGGAATGTTTGGT	255
Db	85	-----GlyAsnValPheSer	89
QY	256	AACGATCGAAACACACAG-----AAACAGTTGAACGCTCTCAGATGACTTTC	303
Db	90	LeuAspArgGluThrArgThrGluAsnAspGlnGluSerGluAspThrArgSerHis	109
QY	304	GGCAGCTCCAGAGATCTTCCGAG-----ATCATGCCCAAGAGCCA	348

Db	110	GlyValLeuLeuGlyArgPheGlnLysAspIleSerGlnGlyLeuLysPheLysGluAla	129
QY	349	GCAGAGAGAGAAATGTTTGAAGAGTGCACAGGCATCTGCGCCACAAAATGATGGG	408
Db	130	TyrGluArgGluValSerLeuLysArg-----	138
QY	409	AAACAGCTGTGCCCCCGGGGAAATCCAAAGTACCTTGGAGAAGATTAAACAAGACATCTGGA	468
Db	139	-----ProLeuGlyAsn---SerProGlyGluArgLeuAsnArg-----	150
QY	469	CCCAAAAGGGGAAACATCGCTGACCCACAGACTCGTGAGAGAAGAGAGCTGTGGTT	528
Db	151	-----LysMetProAspPheGlyGlnValThrVal	160
QY	529	TATGAAGAGATCAGCGACCCCTGAGGAGATGACGAG	564
Db	161	GluGluLysLeuThrProArgGlyGluArgSerGlu	172

Search completed: March 31, 2004, 13:58:04  
Job time : 18 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - protein search, using frame\_plus\_n2p model

Run on: March 31, 2004, 13:40:37 ; Search time 47 Seconds  
(without alignments)  
7733.553 Million cell updates/sec

Title: US-09-975-856-1

Perfect score: 1025

Sequence: 1 ATGACGGAGACACGCCTT.....ATGACGAGTAACCTCCCTCG 576

Scoring table: BLOSUM62

Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 1017041 segs, 315518202 residues

Total number of hits satisfying chosen parameters: 2034082

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%  
Listing first 45 summaries

Command line parameters:

-MODEL=frame\_n2p.model -DEV=xlip  
-Q=/cg2\_1/USPTO.spool\_p/US0975856/runat\_31032004\_132903\_14748/app\_query.fasta\_1.775  
-DB=SPREMBL\_25 -QFMT=fastan -SUFFIX=rspt -MINMATCH=0.1 -LOOPEXT=0 -LOOPEXT=0  
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45  
-DOCALIGN=200 -THR\_SCORE=opt -THR\_MAX=100 -THR\_MIN=0 -ALIGN=15 -MODE=LOCAL  
-OUTFMT=ptlo -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US0975856 -CGN\_1\_1\_86 -runat\_31032004\_132903\_14748 -NCPU=6 -ICPU=3  
-NO\_MMAP -LARGEQUERY -NEG\_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG  
-DEV\_TIMEOUT=120 -WARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

SPREMBL\_25:\*  
1: sp\_archaea:\*  
2: sp\_bacteria:\*  
3: sp\_fungi:\*  
4: sp\_human:\*  
5: sp\_invertebrate:\*  
6: sp\_mammal:\*  
7: sp\_mhc:\*  
8: sp\_organelle:\*  
9: sp\_phase:\*  
10: sp\_plant:\*  
11: sp\_rodent:\*  
12: sp\_virus:\*  
13: sp\_vertebrate:\*  
14: sp\_unclassified:\*  
15: sp\_rvirus:\*  
16: sp\_bacteriap:\*  
17: sp\_archaeap:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	685.5	66.9	223	4	Q9BU88 homo sapien

2	636	62.0	170	4	Q9BRW7	Q9brw7 homo sapien
3	576	56.2	155	4	Q96Q11	Q96q11 homo sapien
4	545.5	53.2	196	4	Q96Q10	Q96q10 homo sapien
5	363	35.4	98	4	Q9Y444	Q9y444 homo sapien
6	255	24.9	64	4	Q8WVZ9	Q8wvz9 homo sapien
7	247.5	24.1	113	4	Q9NZK4	Q9nzk4 homo sapien
8	214.5	20.9	128	11	Q8QZT4	Q8qzt4 mus musculus
9	214.5	20.9	165	11	Q8CSZ3	Q8csz3 mus musculus
10	211.5	20.6	170	11	Q9CPU1	Q9cpul mus musculus
11	160.5	15.7	117	4	Q75101	Q75101 homo sapien
12	137	13.4	39	4	Q8IZH0	Q8izh0 homo sapien
13	137	13.4	39	4	Q8IZG9	Q8izg9 homo sapien
14	137	13.4	39	4	Q8IZG8	Q8izg8 homo sapien
15	137	13.4	39	4	Q8IZG7	Q8izg7 homo sapien
16	137	13.4	39	4	Q8IZG6	Q8izg6 homo sapien
17	137	13.4	39	4	Q8IZG5	Q8izg5 homo sapien
18	124	12.1	39	4	Q8IZH1	Q8izh1 homo sapien
19	113.5	11.1	510	4	Q86TD5	Q86td5 homo sapien
20	112.5	11.0	1111	4	Q60290	Q60290 homo sapien
21	111	10.8	390	4	Q96MN0	Q96mn0 homo sapien
22	110.5	10.8	524	4	Q8N2J5	Q8n2j5 homo sapien
23	110.5	10.8	785	4	Q9ULD5	Q9uld5 homo sapien
24	109.5	10.7	652	11	Q8BFS8	Q8bfs8 mus musculus
25	109	10.6	432	3	Q9HF87	Q9hf87 botrytis ci
26	109	10.6	635	11	Q8QZZ2	Q8qzz2 mus musculus
27	106.5	10.4	782	4	Q8N393	Q8n393 homo sapien
28	106	10.3	579	11	Q9WV10	Q9wv10 mus musculus
29	104.5	10.2	405	4	Q8IW91	Q8iw91 homo sapien
30	104.5	10.2	579	11	Q8C393	Q8c393 mus musculus
31	104	10.1	643	11	Q8BV16	Q8bv16 mus musculus
32	103	10.0	614	11	Q8C887	Q8c887 mus musculus
33	103	10.0	614	11	Q8BPJ3	Q8bpj3 mus musculus
34	103	10.0	621	11	Q8QTC5	Q8qtcs mus musculus
35	102	10.0	511	11	Q8BVH0	Q8bvh0 mus musculus
36	102	10.0	649	4	Q9P215	Q9p215 homo sapien
37	100.5	9.8	182	4	Q9UDV5	Q9udv5 homo sapien
38	100.5	9.8	431	11	Q8C964	Q8c964 mus musculus
39	98.5	9.6	258	17	Q8TPS6	Q8tps6 methanosarc
40	98.5	9.6	428	17	Q8TUH2	Q8tuh2 methanosarc
41	98.5	9.6	428	17	Q8TTV7	Q8ttv7 methanosarc
42	98.5	9.6	428	17	Q8TNA0	Q8tna0 methanosarc
43	98.5	9.6	428	17	Q8TMC5	Q8tmc5 methanosarc
44	98.5	9.6	428	17	Q8TLJ2	Q8tlj2 methanosarc
45	98.5	9.6	428	17	Q8TIL4	Q8til4 methanosarc

#### ALIGNMENTS

RESULT 1  
Q9BU88  
ID Q9BU88 PRELIMINARY; PRT; 223 AA.  
AC Q9BU88;  
DT 01-JUN-2001 (TRENBLrel. 17, Created)  
DT 01-JUN-2001 (TRENBLrel. 17, Last sequence update)  
DT 01-OCT-2003 (TRENBLrel. 25, Last annotation update)  
DE Similar to synovial sarcoma, X breakpoint 2.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Placenta;  
RA Strausberg R.;  
RL Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.  
DR EMBL; BC002818; AA02818.1; -  
DR GO; GO:0005622; C:intracellular; IEA.  
DR GO; GO:0003676; F:nucleic acid binding; IEA.  
DR GO; GO:0006355; P:regulation of transcription, DNA-dependent; IEA.  
DR InterPro; IPR001909; KRAB.  
DR InterPro; IPR003655; KRAB\_related.  
DR Pfam; PF01352; KRAB; 1  
DR SMART; SM00349; KRAB; 1.

```
DR PROSITE; P50806; KRAB RELATED; 1.
SQ SEQUENCE 223 AA; 25173 MW; 2BF9E1FFA4D58094 CRC64;

Alignment Scores:
Pred. No.: 1,15e-62 Length: 223
Score: 685.50 Matches: 137
Percent Similarity: 66.07% Conservatives: 11
Best Local Similarity: 61.16% Mismatches: 26
Query Match: 66.88% Indels: 50
DB: 4 Gaps: 1

US-09-975-856-1 (1-576) x Q9BU88 (1-223)
QY 1 ATGAACGGAGACGACGCTTTGCAAGGAGAGACCCAGGAGATGATCTCAATATATCAGAGAAG 60
Db 1 MetAsnGlyAspAlaPheAlaArgProThrValGlyAlaGlnIleProGluLys 20
QY 61 TTACGAAAGCCCTTCGATGATATTCGCAATATCTTCTTAAGAAAGAGTGGGAAAGATG 120
Db 21 IleGlnLysAlaPheAspAlaLysTyrPheSerLysGluGluTrpGluLysMet 40
QY 121 AATCTCTCGAGAAATCGTCTATGTATGATCAAGCTAAACTATCAGGTCATGACTATAA 180
Db 41 LysAlaSerGluLysIlePheTyrValTyrMetLysArgLysTyrGluAlaMetThrLys 60
QY 181 CTAGGTTTCAAGGTCACCCCTCCACCTTTTCATCGTAGTAACCGGCTGCAGACTTCCAC 240
Db 61 LeuGlyPheLysAlaThrLeuProPheMetCysAsnLysArgAlaGluAspPheGln 80
QY 241 GGAATGATTTGGTACGATCAACACACAGAAATCAGTTGAAAGTCTCGAGACT 300
Db 81 GlyAsnAspLeuAspAsnAspProAsnArgGlyAsnGlnValGluArgProGlnMetThr 100
QY 301 TTCCGAGCCTCCAGAGAAATCTTCCGAGATCATGCCCAAGAGCCAGAGGAAGAA 360
Db 101 PheGlyArgLeuGlnGlyIleSerProLysIleMetProLysLysProAlaGluGly 120
QY 361 AATGTTTGAAGAGTGCAGAGGATCTGGCCCAAAATATGATGGAAACAGCTGTGC 420
Db 121 AsnAspSerGluGluValProGluAlaSerGlyProGlnAsnAspGlyLysGluLys 140
QY 421 CCCCAGGAAATCCAGTACTTGGAGAGATTAAACAGACATCTGGA 468
Db 141 ProGlyLysProThrThrSerGluLysIleHisGluArgSerGlyAsnArgGluAla 160
QY 468 ----- 468
Db 161 GlnGluLysGluGluArgArgGlyThrAlaHisArgTrpSerSerGlnAsnThrHisAsn 180
QY 468 ----- 468
Db 181 IleGlyArgPheSerLeuSerThrSerMetGlyAlaValHisGlyThrProLysThrIle 200
QY 469 -----CCCAAAAGGGGAAACATGCTGGACCCACACAGACTCGGTGAGAGA 513
Db 201 ThrHisAsnArgAspProLysGlyGlyAsnMetProGlyProThrAspCysValArgGly 220
QY 514 AAGCAGCTGG 523
Db 220 uAsnSerTip 223

RESULT 2
Q9BRW7 PRELIMINARY; PRT; 170 AA.
AC Q9BRW7;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Synovial sarcoma, X breakpoint 3.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
RN [1]
RP SEQUENCE FROM N.A.

Q96Q11 PRELIMINARY; PRT; 155 AA.
AC Q96Q11;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE DJ54B20.1.1 (Novel SSX family protein (isoform 1)) (Fragment).
GN DJ54B20.1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
RN [1]
RP SEQUENCE FROM N.A.
```







```

QY 484 CATCCCTGGACCCACAGACTCGTGAGAGAAAGCAGCTGG 523
    |||||
Db 100 nMetProGlyProThrAspCysValArgGluAsnSerTrp 113
    |||||
RESULT 8
ID Q80ZT4 PRELIMINARY; PRT; 128 AA.
AC Q80ZT4;
DT 01-JUN-2003 (TrEMBLrel. 24, Created)
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DE 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Similar to RIKEN cDNA 4930414C09 gene.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC Strausberg R.;
RA Submitted (MAR-2003) to the EMBL/GenBank/DBJ databases.
RL EMBL; BC048441.1; -.
DR GO; GO:0005622; C:intracellular; IEA.
DR GO; GO:0003676; F:nucleic acid binding; IEA.
DR GO; GO:0006355; P:regulation of transcription, DNA-dependent; IEA.
DR InterPro; IPR001909; KRAB.
DR InterPro; IPR003655; KRAB_related.
DR Pfam; PF01352; KRAB; 1.
DR SMART; SM00349; KRAB; 1.
DR PROSITE; PS08005; KRAB; 1.
DR PROSITE; PS08006; KRAB_RELATED; 1.
DR PROSITE; PS08006; KRAB_RELATED; 1.
SQ SEQUENCE 128 AA; 15290 MW; F88614D1CBFF3B70 CRC64;

Alignment Scores:
Pred. No.: 1.5e-13 Length: 128
Score: 214.50 Matches: 56
Percent Similarity: 45.00% Conservative: 25
Best Local Similarity: 31.11% Mismatches: 36
Query Match: 20.93% Indels: 63
DB: 11 Gaps: 5

US-09-975-856-1 (1-576) x Q80ZT4 (1-128)
QY 25 AGGAGACCCAGGATGATGCTCAATATACAGAGAAGTTACGAAAGGCGCTTCGATGATTT 84
    ::: |||||
Db 9 LysAsnProMetGluValLeuTyrGluProLysAsnIleCysLysAlaPheGlnAspIle 28
    ::: |||||
QY 85 GCCAATATCTCTCTAAGAAAGAGTGGGAAAGATGAAATCCCTCGGAGAAATCGTCTAT 144
    ::: |||||
Db 29 SerThrTyrPheSerAspGluGluTyrGlyLysLeuThrGlnTyrGlnLysSerAlaTyr 48
    ::: |||||
QY 145 GTGTATATGAAGTAACTATAGGTCATCACTAACTAGGTTTCAAGGTCCACCTCCCA 204
    |||||
Db 49 ValTyrMetLysArgAsnTyrIleArgMetThrAspLeuGlyValThrValAsnGlnPro 68
    |||||
QY 205 CCTTTCATGCTAGTAAACGGGCTGCAGACTTCCACGGGAATGATTTTGTAAACGATCGA 264
    |||||
Db 69 ValPheMetArg-----GlyLysGluGln 76
    |||||
QY 265 AACACAGGAATCAGTTGAACCTCCTCAGATGACTTTCGGCAGCCTCCAGAGAATCTTC 324
    |||||
Db 77 AspLysGlnSerLeuValGlu-----GlyIle-----GluValHisAsp 83
    |||||
QY 325 CCGAAGATCATGCCCAAGAGCCAGCAGAGAGAAATGTTTGAAGGAAGTCCAGAG 384
    |||||
Db 84 -----GlyIle-----GluValHisAsp 89
    |||||
QY 385 GCATCTGGCCCAACAAATGATGGAAACAGCTGTGCCCCCGGGAATCCAGTACCTTG 444
    |||||
Db 90 Ser----- 90
    |||||
QY 445 GAGAAGATTAAACAGCATCTGCAGCCCAAAAGGGGAAACATGCCTGGACCCACAGACTG 504
    |||||

```

---

```

Db 91 -----AspGluThrSerGlyIleArg-----ValAsnValTrpSerHisArgLeu 105
    |||||
QY 505 CGTGAGAGAAAGCAGCTGGTGTATGAAGATCAGACCTCGAGGAAGATGACGAG 564
    |||||
Db 106 ArgGluArgLysTyrArgValIleTyrGluGluIleSerAspProGluGluGluAsp 125
    |||||
RESULT 9
Q8C5Z3 PRELIMINARY; PRT; 165 AA.
AC Q8C5Z3;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DE 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Hypothetical KRAB box/KRAB-related containing protein.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STEIN=C57BL/6J; TISSUE=Testis;
RX MEDLINE=22354683; PubMed=12466851;
RA The FANTOM Consortium,
RA the RIKEN Genome Exploration Research Group Phase I & II Team;
RT "Analysis of the mouse transcriptome based on functional annotation of
RT 60,770 full-length cDNAs.";
RL Nature 420:563-573(2002).
DR EMBL; AK076879; BAC36519.1; -.
DR GO; GO:0005622; C:intracellular; IEA.
DR GO; GO:0003676; F:nucleic acid binding; IEA.
DR GO; GO:0006355; P:regulation of transcription, DNA-dependent; IEA.
DR InterPro; IPR001909; KRAB.
DR InterPro; IPR003655; KRAB_related.
DR Pfam; PF01352; KRAB; 1.
DR SMART; SM00349; KRAB; 1.
DR PROSITE; PS08005; KRAB; 1.
DR PROSITE; PS08006; KRAB_RELATED; 1.
DR PROSITE; PS08006; KRAB_RELATED; 1.
KW Hypothetical protein.
SQ SEQUENCE 165 AA; 19584 MW; 5FE61134DDB7BAC4 CRC64;

Alignment Scores:
Pred. No.: 1.57e-13 Length: 165
Score: 214.50 Matches: 56
Percent Similarity: 45.00% Conservative: 25
Best Local Similarity: 31.11% Mismatches: 36
Query Match: 20.93% Indels: 63
DB: 11 Gaps: 5

US-09-975-856-1 (1-576) x Q8C5Z3 (1-165)
QY 25 AGGAGACCCAGGATGATGCTCAATATACAGAGAAGTTACGAAAGGCGCTTCGATGATTT 84
    ::: |||||
Db 46 LysAsnProMetGluValLeuTyrGluProLysAsnIleCysLysAlaPheGlnAspIle 65
    ::: |||||
QY 85 GCCAATATCTCTCTAAGAAAGAGTGGGAAAGATGAAATCCCTCGGAGAAATCGTCTAT 144
    ::: |||||
Db 66 SerThrTyrPheSerAspGluGluTyrGlyLysLeuThrGlnTyrGlnLysSerAlaTyr 85
    ::: |||||
QY 145 GTGTATATGAAGTAACTATAGGTCATCACTAACTAGGTTTCAAGGTCCACCTCCCA 204
    |||||
Db 86 ValTyrMetLysArgAsnTyrIleArgMetThrAspLeuGlyValThrValAsnGlnPro 105
    |||||
QY 205 CCTTTCATGCTAGTAAACGGGCTGCAGACTTCCACGGGAATGATTTTGTAAACGATCGA 264
    |||||
Db 106 ValPheMetArg-----GlyLysGluGln 113
    |||||
QY 265 AACACAGGAATCAGTTGAACCTCCTCAGATGACTTTCGGCAGCCTCCAGAGAATCTTC 324
    |||||
Db 114 AspLysGlnSerLeuValGlu-----GlyIle----- 120
    |||||
QY 325 CCGAAGATCATGCCCAAGAGCCAGCAGAGAGAAATGTTTGAAGGAAGTCCAGAG 384
    |||||

```

```

Db 121 -----Glylle-----GluValHisAsp 126
Qy 385 GCATCTGCCCCACAAATGATGGAAACAGCTGTGCCCCCGGAAATCCAAAGTACCTTG 444
Db 127 Ser-----127
Qy 445 GAGAAGATTACAGACATCTGGACCCAAAGGGGAAACATGCTCGACCCACAGACTG 504
Db 128 -----AspGluThrSerGlylleArg-----ValAsnValTrpSerHisArgLeu 142
Qy 505 CGTCAGAGAAAGCAGCTGGTGTATTATCAAGAGATCAGGACCTCGAGGAGATGACGAG 564
Db 143 ArgGluArgLysTyrArgValIleTyrGluGluIleSerAspProGluGluGluAsp 162

RESULT 10
Q9CPU1 PRELIMINARY; PRT; 170 AA.
AC Q9CPU1;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DE 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE 4930414C09Rik protein.
GN 4930414C09Rik.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Testis;
RX MEDLINE=21085660; PubMed=11217851;
RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamataka I.,
RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA Kuehl P., Lewis S., Matsuo J., Nikaido I., Pesole G., Quackenbush J.,
RA Schirni L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA Brownstein M.J., Bull C., Fletcher C., Fujita M., Gariboldi M.,
RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts N.,
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohtsuki S.,
RA Hayaishizaki Y.;
RT "Functional annotation of a full-length mouse cDNA collection.";
RL Nature 409:685-690(2001).
DR EMBL; AK015135; BAB29722.1; -.
DR EMBL; AK006218; BAB24465.1; -.
DR MGD; MGI:1915235; 4930414C09Rik.
DR GO; GO:0005622; C:intracellular; IEA.
DR GO; GO:0003676; P:nucleic acid binding; IEA.
DR GO; GO:0006355; P:regulation of transcription, DNA-dependent; IEA.
DR InterPro; IPR001909; KRAB.
DR InterPro; IPR003655; KRAB_related.
DR Pfam; PF01352; KRAB; 1.
DR SMART; SM00349; KRAB; 1.
DR PROSITE; PS50805; KRAB; 1.
DR PROSITE; PS50806; KRAB_RELATED; 1.
SQ SEQUENCE 170 AA; 19636 MW; F63F702BFDF7227 CRC64;

Alignment Scores:
Pred. No.: 3.24e-13 Length: 170
Score: 211.50 Matches: 58
Percent Similarity: 47.34% Conservative: 22
Best Local Similarity: 34.32% Mismatches: 62
Query Match: 20.63% Indels: 27
DB: 11 Gaps: 5

```

```

US-09-975-856-1 (1-576) x Q9CPU1 (1-170)
Qy 67 AAGCGCTTCGATGATATGCAAAATACTTCTCTAAGAAAGAGTGGGAAAGATGAATCC 126
Db 23 LysAlaPheGlnAspIleSerThrTyrPheSerAspGluGluTrpGlyLysLeuThrGln 42
Qy 127 TCGAGAGAAATCGCTCTATGTGTATATGAAGCTAAACTATGAGGTCTATGATAAAGT 186
Db 43 TrpGlnLysSerAlaTyrValTyrMetLysArgAsnTyrIleArgMetThrAspLeuGly 62
Qy 187 TTCAGAGTCACCTCCACCTTTCATGCTAGTAGTAACACGGGCTGCAGAC-----TTC 237
Db 63 ValThrValAsnGlnProValPheMetArgGlyLysGluGlnAlaLysGlnSerLeuVal 82
Qy 238 CACGGAATGATTTTGTAAACGATCGAAACACCAAGGATTCAGGATTCCTCTCAGATG 297
Db 83 GluGlyLeuVal-----HisAspSerGluAspGluCysPheGluGly 97
Qy 298 ACTTTCGGCAGCTCCAGAGATCTTCCGAGAGATCATGCCCAAGAGACCCAGCAGAGAA 357
Db 98 SerPheGly-----ValThrProIleLysArgMetLysLeu 109
Qy 358 GAAATGGTTTGAAGGAAGTCCAGAGGATCTGGCCCAACAAATGATGGGAAACAGCTG 417
Db 110 ThrSerValThrIleSerPheHisAsnValGluGlySerLeuAlaSerGlyGluAsnAsp 129
Qy 418 TGGCCCGCGGAAATCCAGTACCTTGGAGAGATTAACAAGACATCTGCACCCCAAGG 477
Db 130 Cys-----AsnLeuAlaGluThrGlyGlyLeuGln-- 139
Qy 478 GGAACAATCGCTGGAGCCACACACTCGCTGAGAGAAAGCAGCTGGTGTATGAAAGAG 537
Db 140 ---ValAsnValTrpSerHisArgLeuArgLysTyrArgValIleTyrSerGlu 158
Qy 538 ATCAGCGACCTCGAGGAAGATGACGAG 564
Db 159 IleSerAspThrGluGluGluAsp 167

RESULT 11
Q75101 PRELIMINARY; PRT; 117 AA.
AC Q75101;
DT 01-NOV-1998 (TrEMBLrel. 08, Created)
DT 01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE SSX-HSTT (Fragment).
GN SSX-HSTT.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Sonobe H., Takeuchi T.;
RT "SSX-HSTT".
RL Submitted (MAR-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB012575; BAA32799.1; -.
FT NON TER 1
FT NON TER 117
SQ SEQUENCE 117 AA; 12611 MW; C9F9ADFC0AD4542 CRC64;

Alignment Scores:
Pred. No.: 6.39e-08 Length: 117
Score: 160.50 Matches: 34
Percent Similarity: 64.62% Conservative: 8
Best Local Similarity: 52.31% Mismatches: 16
Query Match: 15.66% Indels: 7
DB: 4 Gaps: 2

US-09-975-856-1 (1-576) x Q75101 (1-117)
Qy 241 GGGAAATGATTTGTAAACGATCGAAACACCAAGGATTCAGGTTGAAGTCTCT-----CAG 294

```

```
Db 53 GlyGlnGlnTyrGlyGlyTyrArgProThrGlnProGlyProGlnProGlnGln 72
QY 295 ATGACTTTCGCGAGCCCTCCAGAGATCTTCCGAGAGATCATGCGAAGATGCGCCAGAGCCAGCGAG 354
Db 73 ArgProTyrGly-----TyrAspGlnIleMetProLysLysProAlaGlu 87
QY 355 GAAGAAATGTTTGAAGAGAGTCCAGAGGCTCTGCGCCACACAAATGATGGAAACAG 414
Db 88 GluGlyAsnAspSerGluGluValProGluAlaSerGlyProGlnAsnAspGlyLysGlu 107
QY 415 CTGTGCCCCCGGGA 429
Db 108 LeuCysProProGly 112

RESULT 12
Q81ZH0 PRELIMINARY; PRT; 39 AA.
AC Q81ZH0;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE SS18/SSX2 fusion protein (Fragment).
GN SS18/SSX2 FUSION.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Wei Y., Sun M., Wang J., Zhu X., Larsson C., Dwight T., Xie Y.,
RA Nilsson G., Larsson O.;
RL Submitted (AUG-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AV138488; AAN39530.1; -.
FT NON_TER 1 39
FT NON_TER 39
SQ SEQUENCE 39 AA; 4336 MW; 8FC179F66C8C7E0D CRC64;

Alignment Scores:
Pred. No.: 1,51e-05 Length: 39
Score: 137.00 Matches: 25
Percent Similarity: 87.88% Conservative: 4
Best Local Similarity: 75.76% Mismatches: 4
Query Match: 13.37% Indels: 0
DB: 4 Gaps: 0

US-09-975-856-1 (1-576) x Q81ZH0 (1-39)

QY 322 TTCCCGAAGATCATGCCAAGAGCCAGAGAGAGAAATGTTTGAAGAGAGTGCCA 381
Db 7 TyrAspGlnIleMetProLysLysProAlaGluGlyAsnAspSerGluGluValPro 26
QY 382 GAGGCATCTGGCCACAAAATGATGGAAACAGCTGTGC 420
Db 27 GluAlaSerGlyProGlnAsnAspGlyLysGluLeuCys 39

RESULT 13
Q81ZG9 PRELIMINARY; PRT; 39 AA.
AC Q81ZG9;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE SS18/SSX2 fusion protein (Fragment).
GN SS18/SSX2 FUSION.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Wei Y., Sun M., Wang J., Zhu X., Larsson C., Dwight T., Xie Y.,
RA Nilsson G., Larsson O.;
RL Submitted (AUG-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AV138488; AAN39530.1; -.
FT NON_TER 1 39
FT NON_TER 39
SQ SEQUENCE 39 AA; 4336 MW; 8FC179F66C8C7E0D CRC64;

Alignment Scores:
Pred. No.: 1,51e-05 Length: 39
Score: 137.00 Matches: 25
Percent Similarity: 87.88% Conservative: 4
Best Local Similarity: 75.76% Mismatches: 4
Query Match: 13.37% Indels: 0
DB: 4 Gaps: 0

US-09-975-856-1 (1-576) x Q81ZG9 (1-39)

QY 322 TTCCCGAAGATCATGCCAAGAGCCAGAGAGAGAAATGTTTGAAGAGAGTGCCA 381
Db 7 TyrAspGlnIleMetProLysLysProAlaGluGlyAsnAspSerGluGluValPro 26
QY 382 GAGGCATCTGGCCACAAAATGATGGAAACAGCTGTGC 420
Db 27 GluAlaSerGlyProGlnAsnAspGlyLysGluLeuCys 39

RESULT 14
Q81ZG8 PRELIMINARY; PRT; 39 AA.
AC Q81ZG8;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE SS18/SSX2 fusion protein (Fragment).
GN SS18/SSX2 FUSION.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Wei Y., Sun M., Wang J., Zhu X., Larsson C., Dwight T., Xie Y.,
RA Nilsson G., Larsson O.;
RL Submitted (AUG-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AV138490; AAN39532.1; -.
FT NON_TER 1 39
FT NON_TER 39
SQ SEQUENCE 39 AA; 4336 MW; 8FC179F66C8C7E0D CRC64;

Alignment Scores:
Pred. No.: 1,51e-05 Length: 39
Score: 137.00 Matches: 25
Percent Similarity: 87.88% Conservative: 4
Best Local Similarity: 75.76% Mismatches: 4
Query Match: 13.37% Indels: 0
DB: 4 Gaps: 0

US-09-975-856-1 (1-576) x Q81ZG8 (1-39)

QY 322 TTCCCGAAGATCATGCCAAGAGCCAGAGAGAGAAATGTTTGAAGAGAGTGCCA 381
Db 7 TyrAspGlnIleMetProLysLysProAlaGluGlyAsnAspSerGluGluValPro 26
QY 382 GAGGCATCTGGCCACAAAATGATGGAAACAGCTGTGC 420
Db 27 GluAlaSerGlyProGlnAsnAspGlyLysGluLeuCys 39

RESULT 15
Q81ZG7 PRELIMINARY; PRT; 39 AA.
AC Q81ZG7;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE SS18/SSX2 fusion protein (Fragment).
GN SS18/SSX2 FUSION.
```

OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Wei Y., Sun M., Wang J., Zhu X., Larsson C., Dwight T., Xie Y.,  
 RA Nilsson G., Larsson O.;  
 RL Submitted (AUG-2002) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AY138491; AAN39533.1; -.  
 FT NON\_TER 1  
 FT NON\_TER 39  
 SQ SEQUENCE 39 AA; 4336 MW; 8FC179F66C8C7E0D CRC64;

Alignment Scores:

Pred. No.:	1.51e-05	Length:	39
Score:	137.00	Matches:	25
Percent Similarity:	87.88%	Conservative:	4
Best Local Similarity:	75.76%	Mismatches:	4
Query Match:	13.37%	Indels:	0
DB:	4	Gaps:	0

US-09-975-856-1 (1-576) x Q8IZG7 (1-39)

Qy	322	TTCCCGAGATCATGCCAGAGCCACAGAGAGAGAAATGGTTTGAAGAAAGTGCCA	381
Db	7	TyrAspGlnIleMetProLysLysProAlaGluGlyAsnAspSerGluGluValPro	26
Qy	382	GAGGCATCTGGCCCAACAAATGATGGGAAACAGCTGTGC	420
Db	27	GluAlaSerGlyProGlnAsnAspGlyLysGluLeuCys	39

Search completed: March 31, 2004, 13:59:50  
 Job time : 50 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 31, 2004, 15:16:19 ; Search time 2745 Seconds  
(without alignments)  
6266.162 Million cell updates/sec

Title: US-09-975-856-1

Perfect score: 576

Sequence: 1 ATGACGGAGACAGCCCTT.....ATGACGAGTAACCTCCCTCG 576

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 27513289 seqs, 14931090276 residues

Total number of hits satisfying chosen parameters: 55026578

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:\*

1: em\_estba:\*  
2: em\_estsum:\*  
3: em\_estin:\*  
4: em\_estnu:\*  
5: em\_estov:\*  
6: em\_estpl:\*  
7: em\_estio:\*  
8: em\_hic:\*  
9: gb\_esti:\*  
10: gb\_est2:\*  
11: gb\_hic:\*  
12: gb\_est3:\*  
13: gb\_est4:\*  
14: gb\_est5:\*  
15: em\_estfun:\*  
16: em\_estom:\*  
17: em\_gss\_hum:\*  
18: em\_gss\_inv:\*  
19: em\_gss\_pln:\*  
20: em\_gss\_vrt:\*  
21: em\_gss\_fun:\*  
22: em\_gss\_man:\*  
23: em\_gss\_mus:\*  
24: em\_gss\_pro:\*  
25: em\_gss\_rod:\*  
26: em\_gss\_pig:\*  
27: em\_gss\_vrl:\*  
28: gb\_gss1:\*  
29: gb\_gss2:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	496.6	86.2	979	12	BM806411
2	495	85.9	836	13	BUI94624
3	495	85.9	887	13	BUI61779
4	491	85.2	863	13	BQ222907
					AGENCOURT
					AGENCOURT
					AGENCOURT
					AGENCOURT
					AGENCOURT
					AGENCOURT
					AGENCOURT
					AGENCOURT
					AGENCOURT
					AGENCOURT
					AGENCOURT
					AGENCOURT
					AGENCOURT
					AGENCOURT
					AGENCOURT
					AGENCOURT
					AGENCOURT
					AGENCOURT
					AGENCOURT
					AGENCOURT
					AGENCOURT
					AGENCOURT
					AGENCOURT
					AGENCOURT
					AGENCOURT
					AGENCOURT
					AGENCOURT
					AGENCOURT
					AGENCOURT
					AGENCOURT
					AGENCOURT
					AGENCOURT
					AGENCOURT
					AGENCOURT
					AGENCOURT
					AGENCOURT
					AGENCOURT
					AGENCOURT
					AGENCOURT
					AGENCOURT
					AGENCOURT
					AGENCOURT
					AGENCOURT
					AGENCOURT
					AGENCOURT
					AGENCOURT
					AGENCOURT
					AGENCOURT
					AGENCOURT
					AGENCOURT
					AGENCOURT
					AGENCOURT
					AGENCOURT
					AGENCOURT
					AGENCOURT
					AGENCOURT
					AGENCOURT
					AGENCOURT
					AGENCOURT
					AGENCOURT
					AGENCOURT
					AGENCOURT
					AGENCOURT
					AGENCOURT
					AGENCOURT
					AGENCOURT
					AGENCOURT
					AGENCOURT
					AGENCOURT
					AGENCOURT
					AGENCOURT
					AGENCOURT
					AGENCOURT
					AGENCOURT
					AGENCOURT
					AGENCOURT
					AGENCOURT
					AGENCOURT
					AGENCOURT
					AGENCOURT
					AGENCOURT
					AGENCOURT
					AGENCOURT
					AGENCOURT
					AGENCOURT
					AGENCOURT
					AGENCOURT
					AGENCOURT
					AGENCOURT
					AGENCOURT
					AGENCOURT
					AGENCOURT
					AGENCOURT
					AGENCOURT
					AGENCOURT
					AGENCOURT
					AGENCOURT
					AGENCOURT
					AGENCOURT
					AGENCOURT
					AGENCOURT
					AGENCOURT
					AGENCOURT
					AGENCOURT
					AGENCOURT
					AGENCOURT
					AGENCOURT
					AGENCOURT
					AGENCOURT
					AGENCOURT
					AGENCOURT
					AGENCOURT
					AGENCOURT
					AGENCOURT
					AGENCOURT
					AGENCOURT
					AGENCOURT
					AGENCOURT
					AGENCOURT
					AGENCOURT
					AGENCOURT
					AGENCOURT
					AGENCOURT
					AGENCOURT
					AGENCOURT
					AGENCOURT
					AGENCOURT
					AGENCOURT
					AGENCOURT
					AGENCOURT
					AGENCOURT
					AGENCOURT
					AGENCOURT
					AGENCOURT
					AGENCOURT
					AGENCOURT
					AGENCOURT
					AGENCOURT
					AGENCOURT
					AGENCOURT
					AGENCOURT
					AGENCOURT
					AGENCOURT
					AGENCOURT
					AGENCOURT
					AGENCOURT
					AGENCOURT
					AGENCOURT
					AGENCOURT
					AGENCOURT
					AGENCOURT
					AGENCOURT
					AGENCOURT
					AGENCOURT
					AGENCOURT
					AGENCOURT
					AGENCOURT
					AGENCOURT
					AGENCOURT
					AGENCOURT
					AGENCOURT
					AGENCOURT
					AGENCOURT
					AGENCOURT
					AGENCOURT
					AGENCOURT
					AGENCOURT
					AGENCOURT
					AGENCOURT
					AGENCOURT
					AGENCOURT
					AGENCOURT
					AGENCOURT
					AGENCOURT
					AGENCOURT
					AGENCOURT
					AGENCOURT
					AGENCOURT
					AGENCOURT
					AGENCOURT
					AGENCOURT
					AGENCOURT
					AGENCOURT
					AGENCOURT
					AGENCOURT
					AGENCOURT
					AGENCOURT
					AGENCOURT
					AGENCOURT
					AGENCOURT

5	488.6	84.8	1001	12	BM471266	AGENCOURT
6	481.6	83.6	781	13	BQ434972	AGENCOURT
7	481.6	83.6	873	13	BUI70242	AGENCOURT
8	481.6	83.6	888	13	BQ432374	AGENCOURT
9	479.2	83.2	1135	10	BE535379	AGENCOURT
10	471.6	81.9	602	14	CB158955	601058532
11	470.8	81.7	950	13	BQ231752	AGENCOURT
12	454.6	78.9	793	12	B1560469	603254435
13	447	77.6	922	13	BQ229996	AGENCOURT
14	423.8	73.6	573	10	BE408883	60103758
15	423.4	73.5	867	10	BE891434	601034893
16	417.6	72.5	627	14	CD767542	AGENCOURT
17	408.8	71.0	589	10	BE410950	60103658
18	402	69.8	943	13	BQ229064	AGENCOURT
19	400.8	69.6	948	12	BG333981	602460335
20	394.4	68.5	630	10	BE408978	60103986
21	394.4	68.5	660	14	CB156144	60103986
22	394.4	68.5	759	10	BE384545	601277946
23	394.4	68.5	798	10	BE410794	601301423
24	388.6	67.5	866	12	BG478295	602523210
25	383.4	66.6	636	10	BE390290	601285178
26	383	66.5	678	10	BF211314	601812555
27	382.4	66.4	757	12	BG104299	602310889
28	377.6	65.6	1053	13	BQ224117	AGENCOURT
29	373	64.8	848	10	BE387586	601274054
30	371.4	64.5	961	13	BUI89304	AGENCOURT
31	370.8	64.4	662	10	BE391023	601285631
32	367.6	63.8	1057	12	BG476029	602521175
33	360.2	62.5	655	12	BE31262	603080529
34	360	62.5	988	10	BE729870	601564905
35	360	62.5	1007	12	BG396977	602433891
36	357.6	62.1	1006	13	BUI89362	AGENCOURT
37	340.6	59.1	533	10	BF184274	601843339
38	339.4	58.9	958	13	BQ232177	AGENCOURT
39	330.6	57.4	623	10	BE896093	601438895
40	326	56.6	367	10	BE866727	601679334
41	324	56.2	927	13	BUI78938	AGENCOURT
42	322	55.9	486	9	AA312651	EST183327
43	322	55.9	896	10	BF212335	601813607
44	319.8	55.5	484	14	CB143725	60143725 K-EST0197
45	317	55.0	820	10	BE777476	601348520

#### ALIGNMENTS

RESULT 1  
BM806411  
LOCUS  
DEFINITION  
BM806411  
5', mRNA Sequence.  
ACCESSION  
BM806411.1  
VERSION  
BM806411.1  
KEYWORDS  
EST.  
SOURCE  
Homo sapiens (human)  
ORGANISM  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE  
1 (bases 1 to 979)  
NIH-MGC <http://mgc.ncl.nih.gov/>  
National Institutes of Health, Mammalian Gene Collection (MGC)  
Unpublished (1999)  
Contact: Robert Strausberg, Ph.D.  
Email: [cgabbs-romail.nih.gov](mailto:cgabbs-romail.nih.gov)  
Tissue Procurement: ATCC/DCTD/DTF  
CDNA Library Preparation: Life Technologies, Inc.  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Agencourt Bioscience Corporation  
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:  
<http://image.llnl.gov>  
Plate: LLNL2258 row: f column: 11  
High quality sequence stop: 622.

BM806411  
AGENCOURT\_6542819 NIH\_MGC\_72 Homo sapiens cDNA clone IMAGE:5548810  
979 bp mRNA linear EST 05-MAR-2002

```
FEATURES
source
  Location/Qualifiers
  1..979
  /organism="Homo sapiens"
  /mol_type="mRNA"
  /db_xref="taxon:9606"
  /clone="IMAGE:5548810"
  /tissue_type="melanotic melanoma"
  /lab_host="DH10B (phage-resistant)"
  /clone_lib="NIH MGC 72"
  /notes="Organ: skin; Vector: pCMV-SPORT6; Site 1: NotI;
  Site 2: SalI; Cloned unidirectionally. Primer: Oligo dr.
  Average insert size 2 kb. Library constructed by Life
  Technologies."
```

ORIGIN

```
Query Match      86.2%; Score 496.6; DB 12; Length 979;
Best Local Similarity 91.5%; Pred. No. 1.8e-131;
Matches 526; Conservative 0; Mismatches 49; Indels 0; Gaps 0;

QY 1 ATGAACGGAGACGACGCTTTGCAAGGAGACCCAGGATGATGCTCAAAATATCAGAGAAG 60
   |||||
Db 65 ATGAACGGAGACGACGCTTTGCAAGGAGACCCAGGATGATGCTCAAAATATCAGAGAAG 124
   |||||
QY 61 TTACGAAAGCCCTTCGATGATATTCGCAAAATATCTTCTTAAGAAAGAGTGGGAAAGATG 120
   |||||
Db 125 ATCCAAAGGCCCTTCGATGATATTCGCAAAATATCTTCTTAAGAAAGAGTGGGAAAGATG 184
   |||||
QY 121 AAATCCTCGAGAAATCGTGTATGTATATGATGAAGTAACTATGAGTCAATGACTAAA 180
   |||||
Db 185 AAAGCCCTCGAGAAATCTTCTATGTGTATATGAAGAAAGATGAGGCTATGACTAAA 244
   |||||
QY 181 CTAGGTTTCAAGGTCACCTCCACCTTCCACCTTTCATGCGTAGTAAACGGGCTGCAGACTTCCAC 240
   |||||
Db 245 CTAGGTTTCAAGGCCACCTTCCACCTTTCATGCGTAGTAAACGGGCTGCAGACTTCCAG 304
   |||||
QY 241 GGAATGATTTGGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAG 300
   |||||
Db 305 GGAATGATTTGGTAAATGATGACCTTAACCGTGGGAATCAGGTTGAACGCTCTCAGATGACT 364
   |||||
QY 301 TTCCGAGACCTCCAGAAATCTTCCGAGATCATGCCGAGAGCCAGCAGAGAGAGAA 360
   |||||
Db 365 TTCCGAGAGCTCCAGGAAATCTCCCGAAGATCATGCCGAGAGCCAGCAGAGAGAGGA 424
   |||||
QY 361 AATGTTTGAAGGAAGTGCAGAGGATCTGGGCCCAAAATATGATGGAACAGCTGTGC 420
   |||||
Db 425 AATGTTTGAAGGAAGTGCAGAGGATCTGGGCCCAAAATATGATGGAACAGCTGTGC 484
   |||||
QY 421 CCCCCGGGAAATCAAGTACCTTGGAGAGATTAACAAAGACATCTGGACCCCAAAAGGGG 480
   |||||
Db 485 CCCCCGGGAAATCAAGTACCTTGGAGAGATTAACAAAGACATCTGGACCCCAAAAGGGG 544
   |||||
QY 481 AAACATGCTGGACCCACAGATGCGTGAGAGAAAGCAGCTGGTGTATGAGAGATC 540
   |||||
Db 545 GAATGCTGGACCCACAGATGCGTGAGAGAAACAGCTGGTGTATGAGAGATC 604
   |||||
QY 541 AGCGACCTTGAGGAAGATGACGAGTAACTCCCTC 575
   |||||
Db 605 AGCGACCTTGAGGAAGATGACGAGTAACTCCCTC 639
   |||||

RESULT 2
BU194624
LOCUS
DEFINITION AGENCOURT_7969283 NIH_MGC_72 Homo sapiens cDNA clone IMAGE:6165903
5', mRNA sequence.
ACCESSION BU194624
VERSION BU194624.1 GI:22708608
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 836)
```

AUTHORS NIH-MGC <http://mgc.nci.nih.gov/>  
National Institutes of Health, Mammalian Gene Collection (MGC)  
Unpublished (1999)  
Contact: Robert Strausberg, Ph.D.  
Email: [cgaps-r@mail.nih.gov](mailto:cgaps-r@mail.nih.gov)  
Tissue Procurement: ATCC/DCTD/BTP  
cDNA Library Preparation: Life Technologies, Inc.  
DNA Sequencing by: Agencourt Bioscience Corporation  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
<http://image.llnl.gov>  
Plate: LLAM13525 row: f column: 16  
High quality sequence stop: 624.

FEATURES  
source  
Location/Qualifiers  
1..836  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="IMAGE:6165903"  
/tissue\_type="melanotic melanoma"  
/lab\_host="DH10B (phage-resistant)"  
/clone\_lib="NIH MGC 72"  
/notes="Organ: skin; Vector: pCMV-SPORT6; Site 1: NotI;  
Site 2: SalI; Cloned unidirectionally. Primer: Oligo dr.  
Average insert size 2 kb. Library constructed by Life  
Technologies."

ORIGIN

```
Query Match      85.9%; Score 495; DB 13; Length 836;
Best Local Similarity 91.3%; Pred. No. 4.9e-131;
Matches 525; Conservative 0; Mismatches 50; Indels 0; Gaps 0;

QY 1 ATGAACGGAGACGACGCTTTGCAAGGAGACCCAGGATGATGCTCAAAATATCAGAGAAG 60
   |||||
Db 74 ATGAACGGAGACGACGCTTTGCAAGGAGACCCAGGATGATGCTCAAAATATCAGAGAAG 133
   |||||
QY 61 TTACGAAAGCCCTTCGATGATATTCGCAAAATATCTTCTTAAGAAAGAGTGGGAAAGATG 120
   |||||
Db 134 ATCCAAAGGCCCTTCGATGATATTCGCAAAATATCTTCTTAAGAAAGAGTGGGAAAGATG 193
   |||||
QY 121 AAATCCTCGAGAAATCGTGTATGATGAAGTAACTATGAGTCAATGACTAAA 180
   |||||
Db 194 AAAGCCCTCAGAGAAATCTTCTATGTGTATGAAGAGAAAGTATGAGGCTATGACTAAA 253
   |||||
QY 181 CTAGGTTTCAAGTCACTCCACCTTTCATGCGTAGTAAACGGGCTGCAGACTTCCAC 240
   |||||
Db 254 CTAGGTTTCAAGGCCACCTTCCACCTTTCATGCGTAGTAAACGGGCTGCAGACTTCCAG 313
   |||||
QY 241 GGAATGATTTGGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAG 300
   |||||
Db 314 GGAATGATTTGGTAAATGATGACCTTAACCGTGGGAATCAGGTTGAACGCTCTCAGATGACT 373
   |||||
QY 301 TTCCGAGACCTCCAGAGATCTTCCGAGAGATCATGCCGAGAGCCAGCAGAGAGAGAA 360
   |||||
Db 374 TTCCGAGAGCTCCAGGAAATCTCCCGAAGATCATGCCGAGAGCCAGCAGAGAGAGGA 433
   |||||
QY 361 AATGTTTGAAGGAAGTGCAGAGGATCTGGGCCCAAAATATGATGGAACAGCTGTGC 420
   |||||
Db 434 AATGATTCGGAGGAAGTGCAGAGGATCTGGGCCCAAAATATGATGGAACAGCTGTGC 493
   |||||
QY 421 CCCCCGGGAAATCCAAAGTACCTTGGAGAGATTAACAAAGACATCTGGACCCCAAAAGGGG 480
   |||||
Db 494 CCCCCGGGAAATCCAAAGTACCTTGGAGAGATTAACAAAGACATCTGGACCCCAAAAGGGG 553
   |||||
QY 481 AAACATGCTGGACCCACAGATGCGTGAGAGAAAGCAGCTGGTGTATGAGAGATC 540
   |||||
Db 554 GAACATGCTGGACCCACAGATGCGTGAGAGAAACAGCTGGTGTATGAGAGATC 613
   |||||
QY 541 AGCGACCTTGAGGAAGATGACGAGTAACTCCCTC 575
   |||||
Db 614 AGCGACCTTGAGGAAGATGACGAGTAACTCCCTC 648
   |||||
```

```

RESULT 3
BU161779
LOCUS
DEFINITION BU161779 887 bp mRNA linear EST 04-SEP-2002
5', mRNA sequence.
ACCESSION BU161779
VERSION BU161779.1 GI:22675689
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 887)
NIH-MGC http://mgc.nci.nih.gov/.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Tissue Procurement: ATCC/DCTP/DTP
cDNA Library Preparation: Life Technologies, Inc.
DNA Sequencing by: The I.M.A.G.E. Consortium (LLNL)
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM13532 row: a column: 22
High quality sequence stop: 739.
FEATURES
source
1..887
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:6168477"
/tissue_type="melanotic melanoma"
/lab_host="DH10B (phage-resistant)"
/clone_lib="NIH-MGC_72"
/notes="Organ: skin; Vector: pCMV-Sport6; Site 1: NotI;
Site 2: SalI; Cloned unidirectionally. Primer: Oligo dT.
Average insert size 2 kb. Library constructed by Life
Technologies."
ORIGIN
Query Match 85.9%; Score 495; DB 13; Length 887;
Best Local Similarity 91.3%; Pred. No. 5.1e-131;
Matches 525; Conservative 0; Mismatches 50; Indels 0; Gaps 0;
QY 1 ATGAACGGAGACGACGCCCTTGCAGAGAGACCCAGGATGATCTCAATATCAGAGAAG 60
Db
75 ATGAACGGAGACGACGCCCTTGCAGAGAGACCCAGGATGATCTCAATATCAGAGAAG 134
QY 61 TTACGAAGGCCCTTCGATGATATGTCCTCAATATCTCTTAAGAAAGAGTGGGAAAGATG 120
Db
135 ATCCAAAGGCCCTTCGATGATATGTCCTCAATATCTCTTAAGAAAGAGTGGGAAAGATG 194
QY 121 AAATCCTCGGAGAAATCTCTATGTTATGAAGCTAAATATGAGGTCTATGACTAA 180
Db
195 AAAGCCTCAGAGAAATCTCTATGTTATGAAGCTAAATATGAGGTCTATGACTAA 254
QY 181 CTAGGTTTCAAGGTCAACCTCCACCTTCATCGCTAGTAAACGGCTCCAGCTTCCAC 240
Db
255 CTAGGTTTCAAGGCCACCTCCACCTTCATGTTGTAATAAACGGGCCGAAGCTCCAG 314
QY 241 GGAATGATTTTGGTAAACGATGAAACACAGGAATCAGGTTGAACGCTCTCAGATGACT 300
Db
315 GGAATGATTTGTAATGATACCTTAACCGTGGGATCAGGTTGAACGCTCTCAGATGACT 374
QY 301 TTCGGCAGCCTCCAGAGATCTCCGAGAGATCATGCCCAAGAGCCAGCAGAGAGAA 360
Db
375 TTCGGCAGCCTCCAGGGAATCTCCCGAAGATCATGCCCAAGAGCCAGCAGAGAGAA 434
QY 361 AATGGTTTGAAGGAAGTCCAGAGGATCTGCCCAACAAATGATGGGAAACAGCTGTGC 420

```

---

```

Db
435 AATGATTCGGGGAAGTCCAGAGAGATCTGGCCCAAAATGATGGGAAAGAGCTGTGC 494
QY
421 CCCCCGGGAATCAAGTACTTTGGAGAGATTAAAGACATCTGGACCCAAAGGGGG 480
Db
495 CCCCCGGGAAACCAACTACTCTGAGAAGATTCAGGAGATCTGGACCCAAAGGGGG 554
QY
481 AAACATCCTCGGACCCACAGACTGCGTGAGAGAAAGCAGCTGGTGTATGAAGATC 540
Db
555 GAACATGCTCGGACCCACAGACTGCGTGAGAGAAACAGCTGGTGTATGAAGATC 614
QY
541 AGGACCTCGGAGAGATGACGATCTCCCTC 575
Db
615 AGGACCTCGGAGAGATGACGATCTCCCTC 649

```

```

RESULT 4
BQ222907
LOCUS
DEFINITION BQ222907 863 bp mRNA linear EST 02-MAY-2002
5', mRNA sequence.
ACCESSION BQ222907
VERSION BQ222907.1 GI:20404307
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 863)
NIH-MGC http://mgc.nci.nih.gov/.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgabbs-r@mail.nih.gov
Tissue Procurement: ATCC/DCTP/DTP
cDNA Library Preparation: Life Technologies, Inc.
DNA Sequencing by: The I.M.A.G.E. Consortium (LLNL)
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM13368 row: h column: 17
High quality sequence stop: 625.
FEATURES
source
1..863
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:6095680"
/tissue_type="melanotic melanoma"
/lab_host="DH10B (phage-resistant)"
/clone_lib="NIH-MGC_72"
/notes="Organ: skin; Vector: pCMV-Sport6; Site 1: NotI;
Site 2: SalI; Cloned unidirectionally. Primer: Oligo dT.
Average insert size 2 kb. Library constructed by Life
Technologies."
ORIGIN
Query Match 85.2%; Score 491; DB 13; Length 863;
Best Local Similarity 91.2%; Pred. No. 7e-130;
Matches 521; Conservative 0; Mismatches 50; Indels 0; Gaps 0;
QY 5 ACGGAGACGACGCCCTTGCAGAGAGACCCAGGATGATCTCAATATCAGAGAAGTAC 64
Db
1 ACGGAGACGACGCCCTTGCAGAGAGACCCAGGATGATCTCAATATCAGAGAAGTAC 60
QY
65 GAAAGCCCTTCGATGATATTCCTCAAGAAAGAGTGGGAAAGATGAAT 124
Db
61 AAGAGCCCTTCGATGATATTCCTCAAGAAAGAGTGGGAAAGATGAAG 120
QY
125 CCTCGGAGAAATCGTCTATGTTATGAAGCTAAATATGAGGTCTATGACTAACTAG 184
Db
121 CCTCAGAGAAATCTCTATGTTATGAAGAGAAAGTATGAGGCTATGACTAACTAG 180

```



QY	185	GTTCACAGCTCACCTCCCTCCACCTTTTCATGCGCTAGTAAACGGGCTCGACACTTCCACGGGA	244
Db	181	GTTCACAGCCACCTCCCTCCACCTTTTCATGTGTAAATAAACGGGCGCAAGACTTCCAGGGGA	240
QY	245	ATGATTTTGGTAAACGATCGAACAACACAGAAATCAGGTTGAAGCTCCTCAGATGACTTTCCG	304
Db	241	ATGATTTTGGATTAATGACCCTTAACCGTGGGAATCAGGTTGAACGCTCCTCAGATGACTTTCCG	300
QY	305	GCAGCCTCCAGAGAATCTTCCCGAAGATCATGCCCAAGAACGCCACGACAGAGAAGAAAATG	364
Db	301	GCAGGCTCCAGGAATCTCCCGGAAGATCATGCCCAAGAACGCCACGACAGAGAAGAAATG	360
QY	365	GTTTGAAGGAAGTGGCAGAGGATCTGGGCCACAAAATGATGGGAAACAGCTGTGSCCCC	424
Db	361	ATTCCGAGGAAGTGCACAGAATCTGGGCCACAAAATGATGGGAAACAGCTGTGSCCCC	420
QY	425	CGGGAAATCCAAAGTACCTTGCAGAGAGATTAAACAGACATCTGGACCCAAAAGGGGGAAAC	484
Db	421	CGGGAAATCCAAATCTACCTTCTGAGAAGATCTCAGAGAGATCTGGACCCAAAAGGGGGAAAC	480
QY	485	ATGCCTTGACCCACAGACTGGCTGAGAGAAAAGCAGCTGGTGGTTTTATGAAGAGATCAGCG	544
Db	481	ATGCCTTGACCCACAGACTGGCTGAGAGAAAACAGCTGGTGGTTTTATGAAGAGATCAGCG	540
QY	545	ACCTGAGGAAGATGACAGTAACTCCCTC	575
Db	541	ACCTGAGGAAGATGACAGTAACTCCCTC	571

[illegible]

Query Match 84.8%; Score 488.6; DB 12; Length 1001;

[illegible]

ORIGIN

Query Match 83.6%; Score 481.6; DB 13; Length 781;  
Best Local Similarity 89.8%; Pred. No. 3.4e-127;  
Matches 517; Conservative 0; Mismatches 59; Indels 0; Gaps 0;

1 ATGACGGAGAGCGCCCTTTCAGAGAGAGCCAGGATGATGCTCAAAATATCAGAGAAG 60  
77 ATGACGGAGAGCGCCCTTTCAGAGAGAGCCAGGATGATGCTCAAAATATCAGAGAAG 136  
61 TTACGAAAGCCCTTCGATGATATGTCACAAATATCTCTTAAGAAAGAGTGGAAAGATG 120  
137 AGAAGCAAGCCCTTGATGATATGTCACAAATATCTCTTAAGAAAGAGTGGAAAGATG 196  
121 AAATCTCTCGAGAAATCGTCTATGTATATGAAGCTAACTATGAGGTCAATGACTTAA 180  
197 AAATCTCTCGAGAAATCGTCTATGTATATGAAGAACTATTAAGGCCATGACTTAA 256  
181 CTAGGTTCTAGAGTCAACCTCCACCTTTCATGCGTAGTAAACGGGTGAGACTTCCAC 240  
257 CTAGGTTCTAGAGTCAACCTCCACCTTTCATGCTGTATTAAGCAAGGCTTCCAG 316  
241 GGAATGATTTTGGTAAACGATCAACACACAGAAATCAGTTTCAACGCTCTCAGTACT 300  
317 GGAATGATTTTGGTAAACGATCAACACACAGAAATCAGTTTCAACGCTCTCAGTACT 376  
301 TTGGGAGCCTCCAGAGATCTTTCCGAGATCATGCCCAAGAGCCAGCAGAGGAGAA 360  
377 TTGGGAGCCTCCAGAGATCATGCCCAAGAGATCATGCCCAAGAGCCAGCAGAGGAGAA 436  
361 AATGTTTGAAGAGTCCAGAGGATCTGGCCCAAAATGATGGGAAACAGCTGTGC 420  
437 AATGTTTGAAGAGTCCAGAGGATCTGGCCCAAAATGATGGGAAACAGCTGTGC 496  
421 CCCCCGGAATCCAAAGTACTTTGGAGAGATTAACAAGACATCTCGACCCCAAAAGGGG 480  
497 CCCCCGGAATCCAAAGTACTTTGGAGAGATTAACAAGACATCTCGACCCCAAAAGGGG 556  
481 AAACATGCTGGACCCACAGACTGCTGGAGAGAAAGCAGCTGCTGATTTATGAAGATC 540  
557 AAACATGCTGGACCCACAGACTGCTGGAGAGAAAGCAGCTGCTGATTTATGAAGATC 616  
541 AGGACCTCTGAGAGATCAGAGTAACTCCCTCG 576  
617 AGGACCTCTGAGAGATCAGAGTAACTCCCTCG 652

RESULT 7  
BUI70242  
LOCUS  
DEFINITION  
AGENCOURT\_7844770 NIH\_MGC\_72 Homo sapiens cDNA clone IMAGE:6051106  
5', mRNA sequence.  
BUI70242  
VERSION  
BUI70242.1 GI:22684226  
KEYWORDS  
EST.  
SOURCE  
Homo sapiens (human)  
ORGANISM  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE  
1 (bases 1 to 873)  
AUTHORS  
NIH-MGC <http://mgc.nci.nih.gov/>.  
TITLE  
National Institutes of Health, Mammalian Gene Collection (MGC)

JOURNAL  
COMMENT

Unpublished (1999)  
Contact: Robert Strausberg, Ph.D.  
Email: [cgapsb@mail.nih.gov](mailto:cgapsb@mail.nih.gov)  
Tissue Procurement: ATCC/DCTD/DTF  
cDNA Library Preparation: Life Technologies, Inc.  
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Agencourt Bioscience Corporation  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
<http://image.llnl.gov>  
Plate: LLAM1304 row: g column: 11  
High quality sequence stop: 696.

FEATURES  
source

Location/Qualifiers  
1..873  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="IMAGE:6051106"  
/tissue\_type="melanotic melanoma"  
/lab\_host="DH10B (phage-resistant)"  
/clone\_lib="NIH\_MGC\_72"  
/note="Organ: skin; Vector: pCMV-SPORT6; Site 1: NotI;  
Site 2: SalI; Cloned unidirectionally. Primer: Oligo dr.  
Average insert size 2 kb. Library constructed by Life  
Technologies."

## ORIGIN

Query Match 83.6%; Score 481.6; DB 13; Length 873;  
Best Local Similarity 89.8%; Pred. No. 3.6e-127;  
Matches 517; Conservative 0; Mismatches 59; Indels 0; Gaps 0;

1 ATGACGGAGAGCGCCCTTTCAGAGAGAGCCAGGATGATGCTCAAAATATCAGAGAAG 60  
80 ATGACGGAGAGCGACACCTTTGCAAGAGAGCCAGGAGATGCTAAAGCATCAGAGAAG 139  
61 TTACGAAAGCCCTTCGATGATATGTCACAAATATCTCTTAAGAAAGAGTGGAAAGATG 120  
140 AGAAGCAAGCCCTTTGATGATATGTCACATATCTTCTTAAGAAAGTGGAAAGATG 199  
121 AAATCTCTCGAGAAATCGTCTATGTATATGAAGCTAACTATGAGGTCAATGACTTAA 180  
200 AAATCTCTCGAGAAATCAGCTATGTATATGAAGAACTATTAAGGCCATGACTTAA 259  
181 CTAGGTTTCAAGTCAACCTCCACCTTTCATGCGTAGTAAACGGGTGAGACTTCCAC 240  
260 CTAGGTTTCAAGTCAACCTCCACCTTTCATGCTGTATTAAGCAAGGCTTCCAG 319  
241 GGAATGATTTTGGTAAACGATCAACACACAGAAATCAGTTTGAACGCTCTCAGATGACT 300  
320 GGAATGATTTTGGTAAATGACCATTAACCCAGATTCAGTTTGAACATCTCAGATGACT 379  
301 TTGGGAGCCTCCAGAGATCTTCCCGAAGATCATGCCCAAGAGCCAGCAGAGGAGAA 360  
380 TTGGGAGCCTCCAGAGATCATGCCGAAGATCATGCCCAAGAGCCAGCAGAGGAGAA 439  
361 AATGTTTGAAGAGTCCAGAGGATCTGGCCCAAAATGATGGGAAACAGCTGTGC 420  
440 AATGTTTGAAGAGTCTCAGAACATCTGGCCCAAAACAGATGGGAAACATCTGCAC 499  
421 CCCCCGGAATCCAAAGTACTTTGGAGAGATTAACAAGACATCTCGACCCCAAAAGGGG 480  
500 CCCCCGGAAGAAAGCAATATTTCTCAGAAGATTAATAAGAGATCTGGACCCCAAGGGG 559  
481 AAACATGCTGGACCCACAGACTGCTGGAGAGAAAGCAGCTGGTGTATGAAGATC 540  
560 AAACATGCTGGACCCACAGACTGCTGGAGAGAAAGCAGCTGGTGTATGAAGATC 619  
541 AGGACCTCTGAGAGATGAGAGTAACTCCCTCG 576  
620 AGGACCTCTGAGAGATGAGAGTAACTCCCTCG 655

## RESULT 8

```
BQ432374
LOCUS      BQ432374      888 bp      mRNA      linear      EST 24-MAY-2002
DEFINITION AGENCOURT 7859139 NIH_MGC_72 Homo sapiens cDNA clone IMAGE:6167925
5', mRNA sequence.
ACCESSION  BQ432374
VERSION    BQ432374.1 GI:21171450
KEYWORDS   EST.
SOURCE     Homo sapiens (human)
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE  NIH-MGC http://mgi.nci.nih.gov/.
            National Institutes of Health, Mammalian Gene Collection (MGC)
            Unpublished (1999)
AUTHORS    Contact: Robert Strausberg, Ph.D.
TITLE      Tissue Procurement: ATCC/DCTD/DTF
JOURNAL    cDNA Library Preparation: Life Technologies, Inc.
COMMENT    DNA Sequencing by: The I.M.A.G.E. Consortium (ILNL)
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/ILNL at:
            http://image.llnl.gov
            Plate: LLM13530 row: j column: 22
            High quality sequence stop: 636.
            Location/Qualifiers
              1..888
                /organism="Homo sapiens"
                /mol_type="mRNA"
                /db_xref="taxon:9606"
                /clone="IMAGE:6167925"
                /tissue_type="melanotic melanoma"
                /lab_host="DH10B (phage-resistant)"
                /clone_lib="NIH MGC 72"
                /note="Organ: skin; Vector: pCMV-SPORT6; Site 1: NotI;
                Site 2: SalI; Cloned unidirectionally. Primer: Oligo dr.
                Average insert size 2 kb. Library constructed by Life
                Technologies."

FEATURES             source
    Query Match      83.6%; Score 481.6; DB 13; Length 888;
    Best Local Similarity 89.8%; Pred. No. 3.6e-127;
    Matches 517; Conservative 0; Mismatches 59; Indels 0; Gaps 0;

QY 1 ATGAACGGAGACGACGCTTTGCAAGGAGACCCAGGATGATGCTCAATATCAGAGAAG 60
DB 79 ATGAACGGAGACGACGCTTTGCAAGGAGACCCAGGATGATGCTCAATATCAGAGAAG 138
QY 61 TTACGAAGGCTTCGATGATATTTGCCAATATCTTCTTAAGAAAGAGTGGGAAAGATG 120
DB 139 AGAAGCAAGGCTTTGATGATATTTGCCACATATCTTCTTAAGAAAGAGTGGGAAAGATG 198
QY 121 AAATCTCTCGGAGAAATCTGCTATGTATGAAGCTAACTATGAGTTCATGACTTAA 180
DB 199 AAATCTCTCGGAGAAATCTGCTATGTATGAAGCTAACTATGAGTTCATGACTTAA 258
QY 181 CTAGGTTTCAAGTTCACCTCCACCTTTCATGCTAGTAAACGGGCTGCAGACTTCCAC 240
DB 259 CTAGGTTTCAAGTTCACCTCCACCTTTCATGCTAGTAAACGGGCTGCAGACTTCCAC 318
QY 241 GGGATGATTTTGGTACGATCGAACCACAGGATCAGGTTGAAACGCTCTCAGATGACT 300
DB 319 GGGATGATTTTGGTACGATCGAACCACAGGATCAGGTTGAAACGCTCTCAGATGACT 378
QY 301 TTCCGAGCTCCAGAGAAATCTTCCGAGATCATGCCCAAGACCCAGCAGAGGAGAA 360
DB 379 TTCCGAGCTCCAGAGAAATCTTCCGAGATCATGCCCAAGACCCAGCAGAGGAGAA 438
QY 361 AATGGTTTGAAGGAAGTGCAGAGGCTATCTGGGCCCAAAATGATGGGAACAGCTGTGC 420
DB 439 AATGATTCGAAGGAGTGTGAGAAGCATCTGGGCCCAAAACGATGGGAAACAACTGCAC 498
```

---

```
QY 421 CCCCCGGGAATCCAACTACCTTGGAGAGATTAAACAAGACATCTGGACCCCAAGGGGG 480
DB 499 CCCCCAGGAAAGCAATATTTCTGAGAGATTAAATAAGAGATCTGGACCCCAAGGGGG 558
QY 481 AAACATGCTCGACCCACAGACTGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGATC 540
DB 559 AAACATGCTCGACCCACAGACTGCTGAGAGAGAGAGAGAGAGAGAGAGAGATC 618
QY 541 AGCGACCTCGAGGAGAGATGACGAGTAATCCCTCG 576
DB 619 AGCGACCTCGAGGAGAGATGACGAGTAATCCCTCG 654

RESULT 9
LOCUS    BE535379
DEFINITION BE535379 NIH_MGC_10 Homo sapiens cDNA clone IMAGE:3445470 5',
mRNA sequence.
ACCESSION BE535379
VERSION    BE535379.1 GI:9764024
KEYWORDS   EST.
SOURCE     Homo sapiens (human)
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE  NIH-MGC http://mgi.nci.nih.gov/.
            National Institutes of Health, Mammalian Gene Collection (MGC)
            Unpublished (1999)
AUTHORS    Contact: Robert Strausberg, Ph.D.
TITLE      Tissue Procurement: ATCC
JOURNAL    cDNA Library Preparation: Life Technologies, Inc.
COMMENT    DNA Sequencing by: Incyte Genomics, Inc.
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/ILNL at:
            http://image.llnl.gov
            Plate: LLM8415 row: o column: 07
            High quality sequence stop: 650.
            Location/Qualifiers
              1..1135
                /organism="Homo sapiens"
                /mol_type="mRNA"
                /db_xref="taxon:9606"
                /clone="IMAGE:3445470"
                /cell_line="MGC36"
                /lab_host="DH10B"
                /clone_lib="NIH MGC 10"
                /note="Organ: cervix; Vector: pCMV-SPORT6; Site 1: NotI;
                Site 2: SalI; Cloned unidirectionally. Primer: Oligo dr.
                Average insert size 1.5 kb. Library prepared by Life
                Technologies."

FEATURES             source
    Query Match      83.2%; Score 479.2; DB 10; Length 1135;
    Best Local Similarity 89.9%; Pred. No. 2e-126;
    Matches 514; Conservative 0; Mismatches 58; Indels 0; Gaps 0;

QY 1 ATGAACGGAGACGACGCTTTGCAAGGAGACCCAGGATGATGCTCAATATCAGAGAAG 60
DB 62 ATGAACGGAGACGACGCTTTGCAAGGAGACCCAGGATGATGCTCAATATCAGAGAAG 121
QY 61 TTACGAAGGCTTCGATGATATTTGCCAATATCTTCTTAAGAAAGAGTGGGAAAGATG 120
DB 122 AGAAGCAAGGCTTTGATGATATTTGCCACATATCTTCTTAAGAAAGAGTGGGAAAGATG 181
QY 121 AAATCTCTCGGAGAAATCTGCTATGTATGAAGCTAACTATGAGTTCATGACTTAA 180
DB 182 AAATCTCTCGGAGAAATCTGCTATGTATGAAGCTAACTATGAGTTCATGACTTAA 241
QY 181 CTAGGTTTCAAGTTCACCTCCACCTTTCATGCTAGTAAACGGGCTGCAGACTTCCAC 240
```

Db 242 CTAGTTTCAAGTCAACCTCCCTCCCTTTCATGTGTAAATAAACAGGCCACAGACTTCCAG 301

Qy 241 GGGAAATGATTTTGGTAAACGATCGAAACACACAGGAATCAGGTTGAACGTCCTCAGATGACT 300

Db 302 GGGAAATGATTTTGAATGACCAATACACCGAGGATTCAGGTTGAACATCTCTCAGATGACT 361

Qy 301 TTGGCAGCTCCAGAGAAATCTTCCGAGAGATCATGCCCAAGAGCCACAGCAGAGGAAGAA 360

Db 362 TTGGCAGCTCCAGAGAAATCATCCGAGAGATCATGCCCAAGAGCCACAGCAGAGGAAGAA 421

Qy 361 AATGTTTGAAGGAGTCCAGAGGATCTTGGCCCAACAAATGATGGGAACAGCTGTGC 420

Db 422 AATGATTCGAAGGAGTGTGAGAGCATCTGGCCCAACAAACGATGGGAACAACTGTGC 481

Qy 421 CCCCCGGAATCAAGTACCTTGGAGAGATTAACAGAGCATCTGGACCCCAAGGGGG 480

Db 482 CCCCCGGAAGCAATATTTCTGAGAGATTAATAGAGATCTGGACCCCAAGGGGG 541

Qy 481 AAACATGCTGGACCCACAGACTCGGTGAGAGAAAGCAGCTGTGTATGAAGAGATC 540

Db 542 AAACATGCTGGACCCACAGACTCGGTGAGAGAAAGCAGCTGTGTATGAAGAGATC 601

Qy 541 AGCAGCCTGAGGAGATGAGGATCACTCC 572

Db 602 AGCAGCCTGAGGAGATGAGGATCACTCC 633

## RESULT 10

CB158955

LOCUS

K-EST0218399 L18POOL1nl Homo sapiens cDNA clone L18POOL1nl-32-F07

5', mRNA sequence.

ACCESSION

CB158955

VERSION

CB158955.1

KEYWORDS

EST.

SOURCE

Homo sapiens (human)

ORGANISM

Homo sapiens

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

Genome Research Center

Korea Research Institute of Bioscience &amp; Biotechnology

52 Boeun-dong Yuseong-gu, Daejeon 305-333, South Korea

Tel: +82-42-860-4470

Fax: +82-42-860-4409

Email: yongsung@mail.kribb.re.kr

Plate: 32 row: F column: 07

High quality sequence stop: 602.

Location/Qualifiers

1. .602

/organism="Homo sapiens"

/mol\_type="mRNA"

/db\_xref="taxon:9606"

/clone="L18POOL1nl-32-F07"

/cell\_line="SNU-354+Cho-CK+Choi-CK+HLK-3"

/lab\_host="Top10F"

/clone\_lib="L18POOL1nl"

/notes="Organ: Liver; Vector: pTVT3-Pac; Site: 1: EcoRI;

Site: 2: NotI; The library was contributed by the Soares

laboratory and it was constructed as described by Bonaldo,

M.F., Lennon, G. and Soares, M.B. (1996), Genome Research

6(9): 791-806. RNA was prepared from harvested cell

culture."

culture."

culture."

culture."

culture."

culture."

culture."

Matches 501; Conservative 0; Mismatches 49; Indels 0; Gaps 0;

Qy 1 ATGAACGAGACGACGCTTTGCAAGGAGAGCCAGGAGATGCTCAATATCAGAGAAG 60

Db 53 ATGAACGAGACGACGCTTTGCAAGGAGAGCCAGGAGATGCTCAATATCAGAGAAG 112

Qy 61 TTACGAAGGCTTCGATGATATTCGCAAAATCTCTCTAAGAAAGAGTGGGAAAAGATG 120

Db 113 ATCCAAAGGCTTCGATGATATTCGCAAAATCTCTCTAAGAAAGAGTGGGAAAAGATG 172

Qy 121 AATCTCTCGAGAAATCGTCTATGTATATGAAGCTAAACTATGAGGTCTACACTAAA 180

Db 173 AAGCCTCGAGAAATCTCTATGTATATGAAGAAAGATGAGGCTATGACTAAA 232

Qy 181 CTAGTTTCAAGTCAACCTCCACCTTTCATGTATGAAGCTAAACTATGAGGTCTACACTAAA 240

Db 233 CTAGTTTCAAGTCAACCTCCACCTTTCATGTATGAAGCTAAACTATGAGGTCTACACTAAA 292

Qy 241 GGGAAATGATTTTGTACAGATCGAAACCAACAGGATCAGGTTGAACGCTCTCAGATGACT 300

Db 293 GGGAAATGATTTTGTACAGATCGAAACCAACAGGATCAGGTTGAACGCTCTCAGATGACT 352

Qy 301 TTGGCAGCTCCAGAGAAATCTTCCGAGAGATCATGCCCAAGAGCCACAGAGGAAGAA 360

Db 353 TTGGCAGCTCCAGAGAAATCTTCCGAGAGATCATGCCCAAGAGCCACAGAGGAAGAA 412

Qy 361 AATGTTTGAAGGAGTCCAGAGGATCTTCCGAGAGATCATGCCCAAGAGCCACAGAGGAAGAA 420

Db 413 AATGATTCGAGGAGAGTCCAGAGGATCTTCCGAGAGATCATGCCCAAGAGCCACAGAGGAAGAA 472

Qy 421 CCCCCGGAATCCAAAGTACCTTGGAGAGATTAACAGAGCATCTGGACCCCAAGGGGG 480

Db 473 CCCCCGGAATCCAAAGTACCTTGGAGAGATTAACAGAGCATCTGGACCCCAAGGGGG 532

Qy 481 AAACATGCTGGACCCACAGACTCGGTGAGAGAAAGCAGCTGTGTATGAAGAGATC 540

Db 533 GAACATGCTGGACCCACAGACTCGGTGAGAGAAAGCAGCTGTGTATGAAGAGATC 592

Qy 541 AGCAGCCTG 550

Db 593 AGCAGCCTG 602

## RESULT 11

BQ231752

LOCUS

DEFINITION

5', mRNA sequence.

ACCESSION

BQ231752

VERSION

BQ231752.1

KEYWORDS

EST.

SOURCE

Homo sapiens (human)

ORGANISM

Homo sapiens

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

NIH-MGC http://mgi.nci.nih.gov/

National Institutes of Health, Mammalian Gene Collection (MGC)

Unpublished (1999)

Contact: Robert Strausberg, Ph.D.

Email: cgapbs@mail.nih.gov

Tissue Procurement: ATCC/DCTD/DTF

cDNA Library Preparation: Life Technologies, Inc.

cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

DNA Sequencing by: Agencourt Bioscience Corporation

Clone distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at:

http://image.llnl.gov

Plate: LLNL3343 row: i column: 08

High quality sequence stop: 532.

## FEATURES

source

Location/Qualifiers

1. .950

/organism="Homo sapiens"

```

/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:6066127"
/tissue_type="melanotic melanoma"
/lab_host="DH10B (phage-resistant)"
/clone_lib="NIH_MGC_72"
/note="Organ: skin; Vector: pCMV-SPORT6; Site:1: NotI; Site 2: SalI; Cloned unidirectionally. Primer: Oligo dt. Average insert size 2 kb. Library constructed by Life Technologies."

ORIGIN
      Query Match      81.7%; Score 470.8; DB 13; Length 950;
      Best Local Similarity 89.9%; Pred. No. 4.8e-124;
      Matches 516; Conservative 0; Mismatches 57; Indels 1; Gaps 1;

QY 1 ATGAACGGAGACGACGCGCTTTGCAAGAGAGACCCAGGAGATGATCTCAATATACAGAGAAG 60
DB 74 ATGAACGGAGACGACGCGCTTTGCAAGAGAGACCCAGGTTGGTGTCTCAATACAGAGAAG 133
QY 61 TTACGAAGGCGCTTCGATGATATTGCGCAATATCTTCTAAGAAGAGTGGGAAAGATG 120
DB 134 ATCCAAAGGCGCTTCGATGATATTGCCAAATACTTCTTAAGGAAGAGTGGGAAAGATG 193
QY 121 AATCTCTCGGAGAAATCGTCTATGTCTATATGAAGCTTAACHTATGAGTCACTACATAA 180
DB 194 AAGCCCTCAGAGAAATCTTCTATGTATATGAAGAGAAAGTATGAGGCTATGACTAAA 253
QY 181 CTAGGTTTCAAGGTCACCCCTCCCAACCTTTTCATGCGTAGTAAACGGGCTCGAGACTTCCAC 240
DB 254 CTAGGTTTCAAGGCCACCCCTCCCAACCTTTTCATGTGTATATAGCGGGCGGAGACTTCAG 313
QY 241 GGGNATGATTTTGGTAACGATTCGAAACACAGGAATCAGGTTGAACGTCCTCAGATCACT 300
DB 314 GGGNATCAGCTGGATAATATGACCCCTTAACCGTGGGAATCAGGTTGAACGTCCTCAGATCACT 373
QY 301 TTCGGCAGCTCCAGAGAAATCTTCCCAAGATCATGCCAGAGAGCCAGCAGAGAGAGAA 360
DB 374 TTCGGCAGGCTTCAGGGAATCTCCCCGAAGATCATGCCAAGAGCCAGCAGAGAGAGGA 433
QY 361 AATGGTTTGAAGGAAGTGCAGAGGCCATCTGGCCCACAAAATGATGGAAACAGCTGTGC 420
DB 434 AATGATTCGGAGGAAGTGCCAGAAGCATCTGGCCCCACAANTGATGGAAAGAGCTGTGC 493
QY 421 CCCCAGGGAATCCAAGTACCTTGGAGAGATTAACAAGACATCTGACCCAAAAGGGGG 480
DB 494 CCCCAGGGAATAACCACTACTCTTGAGAAGATTCCAGAGAGATCTGACCCAAAAGGGGG 553
QY 481 AAACAT-GCCTGGACCCACAGACTGCGTGTAGAGAGAAACAGCTGCTGCTTTATGAGAGAT 539
DB 554 GACATGGCCCTGACCCACAGACTGCGTGTAGAGAGAAACAGCTGCTGCTTTATGAGAGAT 613
QY 540 CAGGACCCCTGAGGAGATCAGCAGTAACCTCCCC 573
DB 614 CAGCGACCTCTGGAGGAGATCAGCAGTAACCTCCCC 647

```

RESULT 12  
 BI560469  
 LOCUS  
 DEFINITION 793 bp mRNA linear EST 05-SEP-2001  
 603284435F1 NIH\_MGC\_97 Homo sapiens cDNA clone IMAGE:5296860 5',  
 mRNA sequence.  
 ACCESSION BI560469  
 VERSION BI560469  
 KEYWORDS EST. GI:15447770  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 REFERENCE 1 (bases 1 to 793)  
 NIH-MGC <http://mgc.nci.nih.gov/>.  
 AUTHORS National Institutes of Health, Mammalian Gene Collection (MGC)  
 TITLE Unpublished (1999)  
 JOURNAL

Contact: Robert Strausberg, Ph.D.  
Email: [cgabbs@mail.nih.gov](mailto:cgabbs@mail.nih.gov)  
Tissue Procurement: Miklos Palkovits, M.D., Ph.D.  
cDNA library preparation: Michael J. Brownstein (NHGRI), Shitaki  
Toshiyuki and Piero Carninci (RIKEN)  
cDNA library Arrayed by: The I.M.A.G.E. Consortium (LNLNL)  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LINL at:

http://image.llnl.gov  
Plate: LLAM11751 row: e column: 09  
High quality sequence stop: 790.

FEATURES source

```

1. .793
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:5296890"
/lab_host="DH10B"
/clone_lib="NIH MGC 97"

```

(containing NIH MG-1) / notes="Organ: testis; Vector: pBluescriptR (modified pBluescript KS+); Site 1: BamHI; Site 2: SalI-XhoI (gtcgag); Oligo-dT primed using primer 5'-TTTTTTTTTTTTTNN-3', size-selected for average insert size 2.2 kb and normalized to ROT 5. This is a primary library enriched for full-length clones and constructed using the Cap-trapper method (Carninci, in preparation). Library constructed by M. Brownstein (NTM/ANHRJ, National Institutes of Health). Note: this is a NIH MG1 Library."

## ORIGIN

Query Match 78.9%; Score 454.6; DB 12; Length 793;  
Best Local Similarity 90.1%; Pred. No. 2e-119;  
Matches 520; Conservative 0; Mismatches 54; Indels 3; Gaps 3;

1	ATGAA	CGGAGACGACGCGCTTTGCAA	-GGAGACCCAGGGATGATGCTCAAAATATCAGAGAA	59
143	ATGAA	CGGAGACGACGCGCTTTGCAAGGGAGACCCACGGTTGGTGTCTCAAAATATCAGAGAA	202	
60	GTTACGAAAGGCGCTTCGATGATATATGCCAAATACTTCTCTAAGAAAGAGTGGGAAAGAT	119		
203	GATCCAAAAGGCGCTTCGATGATATATGCCAAATACTTCTCTAAGAAAGAGTGGGAAAGAT	262		
120	GAATCCTCGGAGAAATCGTCTATGTGTATATGAAGCTAAACTATCAGGCTCATGACTAA	179		
263	GAAAGCCTCGGAGAAATCTTCTATGTGTATATGAGAGAAAGTATGAGGCTATGACTAA	322		
180	ACTAGTTCCTCAAGGTCACCCCTCCACCTTCATCGGTAGTAAACGGGCTGCAGACTTCCA	239		
323	ACTAGGTTTCAAGGCCACCCCTCCACCTTCATGTGTAATAAACGGCGCGAAGACTTCCA	382		
240	CGGGATGATTTTGGTTAAAGATCGAGAAACACACAGGAATCAGTTGAACGTCCTCAGATGAC	299		
383	-GGGAATGATTTGGATTAATGACCCCTAACCGTGGGAATCAGTTGAACGTCCTCAGATGAC	441		
300	TTTCGGCAGCGCTCCAGAGAAATCTTCCCGAAGATCATGCCCCAAGAAGCCAGCAGAGAAGA	359		
442	TTTCGGCAGCGCTCCAGGGAATCTCCCGAAGATCATGCCAAGAGCCAGCAGAGAAGG	501		
360	AAATGTTTTGAAGGAAGTCCAGAGGCATCTCGGCCACAAAAATGATGGGAAACAGCTGTG	419		
502	AAATGATTTCGGAGGAAGTSCCAGAAGCATCTGGCCACAAAAATGATGGGAAAGAGCTGTG	561		
420	CCCCCGGGAATCCNAGTACCTTGGAGAGAGATTAAAGACATCTGGAGCCCAAAAGGGG	479		
562	CCCCCGGGAACCAACTACCTCTGAGAAGATTCAAGAGAGATCTGGAGCCCAAAAGGGG	621		
480	GAACATGCTCGACCCACAGACTCGGTGAGAGAAAGCAGCTGGTGTATTATGAA-GAGA	538		
622	GGAAATGCTCGACCCCGAGACTCGGTGAGAGAAACAGCTGGTGTATTATGAAACGAGA	681		
539	TCAGGACCGCTCAGGAAGATGACGAGTAACCTCCCTC	575		

682 TCAGCGCCTGAGGAAGATGACGAGTACTCCCTC 718

361 AATGTTTGAAGAGTGCAGAGGATCTTGGCCCAAAATGATGGGAAACAGCTGTGC 42

404 AATGATTCGAAGGAGTGTGAGAAGCATCTGGCCCAAAACGATGGGAAACAACTGCAC 463

421 CCCCCGGGAAATCAAGTACCTTGGAGAGATTAAACAAGACATCTCGGCCCAAAA-GGGG 479

464 CCCCCGAAAGCAATTAATCTGAGAGATTAAAGAGATCTGGACCCAAAAGGGGG 523

480 GAAACATGCTGGACCCACAGACTGCGTGAGAGAAAGCAGCTGGTGG-TTTATGAAGAGA 538

524 GAAACATGCTGGACCCACAGACTGCGTGAGAGATTAAAGAGATCTGGACCCAAAAGGGGG 523

539 TCAG-CGACCCCTCAGAGATGACGAGTAACCTCCCTCG 576

584 TCAGCCGACCTCAGAGATGACGAGTAACCTCCCTCG 622

BE408883 573 bp mRNA linear EST 21-JUL-2000  
601303758F1 NIH\_MGC\_21 Homo sapiens cDNA clone IMAGE:3637982 5',  
mRNA sequence.

BE408883  
BE408883.1 GI:9345333  
EST.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
NIH-MGC http://mgi.nci.nih.gov/  
National Institutes of Health, Mammalian Gene Collection (MGC)  
Unpublished (1999)  
Contact: Robert Strausberg, Ph.D.  
Email: cgapbs-r@mail.nih.gov  
Tissue Procurement: ATCC  
cDNA Library Preparation: Ling Hong/Rubin Laboratory  
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov  
Plate: LLM339 row: d column: 15  
High quality sequence stop: 571.  
Location/Qualifiers  
1..573  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="IMAGE:3637982"  
/tissue\_type="choriocarcinoma"  
/lab\_host="DH10B (phage-resistant)"  
/clone\_lib="NIH MGC 21"  
/note="Organ: placenta; Vector: pOTB7; Site: 1: XhoI;  
Site 2: EcoRI; cDNA made by oligo-dT priming.  
Directionally cloned into EcoRI/XhoI sites using the  
following 5' adaptor: GCGAGAG(G). Size-selected >500bp  
for average insert size 1.8kb. Library constructed by  
Ling Hong in the laboratory of Gerald M. Rubin (University  
of California, Berkeley) using ZAP-cDNA synthesis kit  
(Stratagene) and Superscript II RT (Life Technologies)."

Query Match 73.6%; Score 423.8; DB 10; Length 573;  
Best Local Similarity 88.9%; Pred. No. 1.2e-110;  
Matches 458; Conservative 0; Mismatches 57; Indels 0; Gaps 0;

1 ATGAACGGAGACGACGCTTTGCAAGGAGACCCAGGATGATGCTCAAAATCAGAGAAG 60  
59 ATGAACGGAGACGACGCTTTGCAAGGAGACCCAGGATGATGCTCAAAATCAGAGAAG 118  
61 TTACGAAGGCTTCGATGATGATGCTCAAAATCCTTCTTAAGAAAGAGTGGGAAAGATG 120

361 AATGTTTGAAGAGTGCAGAGGATCTTGGCCCAAAATGATGGGAAACAGCTGTGC 42

404 AATGATTCGAAGGAGTGTGAGAAGCATCTGGCCCAAAACGATGGGAAACAACTGCAC 463

421 CCCCCGGGAAATCAAGTACCTTGGAGAGATTAAACAAGACATCTCGGCCCAAAA-GGGG 479

464 CCCCCGAAAGCAATTAATCTGAGAGATTAAAGAGATCTGGACCCAAAAGGGGG 523

480 GAAACATGCTGGACCCACAGACTGCGTGAGAGAAAGCAGCTGGTGG-TTTATGAAGAGA 538

524 GAAACATGCTGGACCCACAGACTGCGTGAGAGATTAAAGAGATCTGGACCCAAAAGGGGG 523

539 TCAG-CGACCCCTCAGAGATGACGAGTAACCTCCCTCG 576

584 TCAGCCGACCTCAGAGATGACGAGTAACCTCCCTCG 622

BE408883 573 bp mRNA linear EST 21-JUL-2000  
601303758F1 NIH\_MGC\_21 Homo sapiens cDNA clone IMAGE:3637982 5',  
mRNA sequence.

BE408883  
BE408883.1 GI:9345333  
EST.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
NIH-MGC http://mgi.nci.nih.gov/  
National Institutes of Health, Mammalian Gene Collection (MGC)  
Unpublished (1999)  
Contact: Robert Strausberg, Ph.D.  
Email: cgapbs-r@mail.nih.gov  
Tissue Procurement: ATCC  
cDNA Library Preparation: Ling Hong/Rubin Laboratory  
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov  
Plate: LLM339 row: d column: 15  
High quality sequence stop: 571.  
Location/Qualifiers  
1..573  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="IMAGE:3637982"  
/tissue\_type="choriocarcinoma"  
/lab\_host="DH10B (phage-resistant)"  
/clone\_lib="NIH MGC 21"  
/note="Organ: placenta; Vector: pOTB7; Site: 1: XhoI;  
Site 2: EcoRI; cDNA made by oligo-dT priming.  
Directionally cloned into EcoRI/XhoI sites using the  
following 5' adaptor: GCGAGAG(G). Size-selected >500bp  
for average insert size 1.8kb. Library constructed by  
Ling Hong in the laboratory of Gerald M. Rubin (University  
of California, Berkeley) using ZAP-cDNA synthesis kit  
(Stratagene) and Superscript II RT (Life Technologies)."

Query Match 73.6%; Score 423.8; DB 10; Length 573;  
Best Local Similarity 88.9%; Pred. No. 1.2e-110;  
Matches 458; Conservative 0; Mismatches 57; Indels 0; Gaps 0;

1 ATGAACGGAGACGACGCTTTGCAAGGAGACCCAGGATGATGCTCAAAATCAGAGAAG 60  
59 ATGAACGGAGACGACGCTTTGCAAGGAGACCCAGGATGATGCTCAAAATCAGAGAAG 118  
61 TTACGAAGGCTTCGATGATGATGCTCAAAATCCTTCTTAAGAAAGAGTGGGAAAGATG 120

```

Db      119 AGAAGCAAGCCCTTTGATGATATTGCCACATATTTCTTAAGAAAGAGTGGAAAAGATG 178
Qy      121 AATATCTCGAGAAATCTCTATGTGTATATGAAGTAACTATGAGGTCTATGACTAAA 180
Db      179 AATATCTCGAGAAATCAGCTATGTGTATATGAAGAAACTATAAGGCCATGACTAAA 238
Qy      181 CTAGGTTTCAAGTCAACCTCCACCTTTTCATGCGTAGTAAACGGGCTCAGACTCCAC 240
Db      239 CTAGGTTTCAAGTCAACCTCCACCTTTTCATGCGTAGTAAACGGGCTCAGACTCCAC 298
Qy      241 GGAATGATTTTGTAAATCATCGAAACCAAGGAATCAGGTTGAAGCTCCTCAGATGACT 300
Db      299 GGAATGATTTTGTAAATCATCGAAACCAAGGAATCAGGTTGAAGCTCCTCAGATGACT 358
Qy      301 TTCCGAGCCTCCAGAGATCTTCCGAGATCATGCCAGATCATGCCAGATCATGCCAGAT 360
Db      359 TTCCGAGCCTCCAGAGATCATGCCAGATCATGCCAGATCATGCCAGATCATGCCAGAT 418
Qy      361 AATGTTTGAAGAAAGTGCAGAGGATCTGGCCCAACAAATGATGGGAACAGCTGTGC 420
Db      419 AATGATTCGAAGGAGTGTCAAGAGATCTGGCCCAACAAATGATGGGAACAGCTGTGC 478
Qy      421 CCCCAGGAAATCCAGTACCTTGAGAGATTAACAAGATCTGGACCCCAAAAGGGGG 480
Db      479 CCCCAGGAAAGCAAAATATTCTGAGAAGATTAATAAGAGATCTGGACCCCAAAAGGGGG 538
Qy      481 AAACATGCTGGACCCACAGATCGGTGAGAGAA 515
Db      539 AAACATGCTGGACCCACAGATCGGTGAGAGAA 573

```

```

RESULT 15
BE891434
LOCUS   601434893F1 NIH_MGC_72 Homo sapiens cDNA clone IMAGE:3919991 5',
        mRNA sequence.
ACCESSION BE891434
VERSION   BE891434.1 GI:10350764
KEYWORDS EST.
SOURCE    Homo sapiens (human)
ORGANISM Homo sapiens
          Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
          Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 867)
          NIH-MGC http://mgc.nci.nih.gov/.
          National Institutes of Health, Mammalian Gene Collection (MGC)
          Unpublished (1999)
          Contact: Robert Strausberg, Ph.D.
          Email: cgabbs-remail.nih.gov
          Tissue Procurement: ATCC/DCRN/Drp
          cDNA Library Preparation: Life Technologies, Inc.
          cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
          DNA Sequencing by: Incyte Genomics, Inc.
          Clone distribution: MGC clone distribution information can be
          found through the I.M.A.G.E. Consortium/LLNL at:
          http://image.llnl.gov
          Plate: L1AM9750 row: j column: 24
          High quality sequence stop: 641.
          Location/Qualifiers
            1..867
              /organism="Homo sapiens"
              /mol_type="mRNA"
              /db_xref="taxon:9606"
              /clone="IMAGE:3919991"
              /tissue_type="melanotic melanoma"
              /lab_host="DH10B (phage-resistant)"
              /clone_lib="NIH MGC 72"
              /notes="Organ; skin; Vector: pCMV-SPORT6; Site 1: NotI;
              Site 2: SalI; Cloned unidirectionally. Primer: Oligo dT.
              Average insert size 2 kb. Library constructed by Life
              Technologies."

```

FEATURES  
source

```

Query Match      73.5%; Score 423.4; DB 10; Length 867;
Best Local Similarity 89.4%; Pred. No. 1.9e-110;
Matches 512; Conservative 0; Mismatches 56; Indels 5; Gaps 5;

Qy      1 ATGAACGAGACGACGCGCTTTGCAAGGAGAGCCAGGGATGATGCTCAAAATATCAGAGAAG 60
Db      50 ATGAACGAGACGACGCGCTTTGCAAGGAGAGCCAGGGATGATGCTCAAAATATCAGAGAAG 109
Qy      61 TTACGAAAGGCGCTTCGATGATATTGCCAAATACCTTCTTAAGAAAGAGTGGAAAAGATG 120
Db      110 ATCCAAAGGCGCTTCGATGATATTGCCAAATACCTTCTTAAGGAGAGTGGAAAAGATG 169
Qy      121 AAATCTCGGAGAAATCGTCTATGTGTATATGAAGCTAAACTATGAGGTCTATGACTAAA 180
Db      170 AAAGCTCAGAGAAATCTTCTATGTGTATATGAAGAAAGTATGAGGCTATGACTAAA 229
Qy      181 CTAGGTTTCAAGTCAACCTCCACCTTTTCATGCGTAGTAAACGGGCTGAGACTTCCAC 240
Db      230 CTAGGTTTCAAGGCGCGCTCCACCTTTCATGTGTATTAATAACGGGCCGAAAGACTTCCA- 288
Qy      241 GGAATGATTTTGTAAACGATCGAAACCAAGGAATCAGGTTGAAGCTGAACTCTCAGATGACT 300
Db      289 GGAATGATTTTGTAAATGACCTTAACCGTGGGAAATCAGGTTGAACGCTCTCAGATGACT 348
Qy      301 TTCCGAGCCTCCAGAGATCTTCCCGAAGATCATGCCCAAGAGCCAGAGAGAGAAAGAA 360
Db      349 TTCCGAGGCTCCAGAGAAATCTCCCGAAGATCATGCCCAAGAGCCAGAGAGAGAAAGAA 408
Qy      361 AATGTTTGAAGGAGTCCAGAGGATCTGGCCCAACAAATGATGGGAACAGCTGTGC 420
Db      409 AATGATTCGAGGAGTGGCAGAGATCTGGCCCAACAAATGATGGGAACAGCTGTGC 468
Qy      421 CCCCAGGAAATCCAGTACCTTTGGAGAAGATTAAACAAGACATCTGGACCC-AAAAAGGG 479
Db      469 CCCCAGGAAA-ACNACTACTCTGAGAAGATTCAGAGAGATCTTGACCCCAAAAGGGG 527
Qy      480 GAAACATGCTCGAACCCACAGACTGCGTGTAGAGAAAGCAGCTGGTGTATGAAGAGAT 539
Db      528 GAAACATGCTCGAACCCACAGACTGCGTGTAGAGAAAGCAG- TGGTGTATGACGAGAT 586
Qy      540 CAGCAGCCTTCAGGAAAGATGACGAGTAACTCCC 572
Db      587 CAGCGA-CTGAGGAAAGATGACGAGTTAACTCC 618

```

Search completed: March 31, 2004, 17:25:05  
Job time : 2750 secs

ORIGIN